NETWORK - WIDE TRANSIT SYSTEM SCHEDULING USING GENETIC ALGORITHMS

by

BANDARU SRINIVAS

CE
1995
M
SRI
NET

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

DECEMBER, 1995

NETWORK-WIDE TRANSIT SYSTEM SCHEDULING USING GENETIC ALGORITHMS

A Thesis Submitted
in Partial Fulfilment of the Requirements
for the Degree of
Master of Technology

by
Bandaru Srinivas

to the
DEPARTMENT OF CIVIL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
December, 1995

30 APR 1996
CENTRAL LIBRARY
I.I.T. KANPUR
Acc. No. A. 121391

A121391

CE-1995-M-SRI-NET

CERTIFICATE

It is certified that the work contained in the thesis entitled **Network-wide Transit System Scheduling using Genetic algorithms**, by **Bandaru Srinivas**, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Partha Chakraborty

Signature of Supervisor
Name : *Dr. Partha Chakroborty*
Assistant Professor
Department of Civil Engineering
I.I.T. Kanpur

December, 1995

POST GRADUATE OFFICE
I.I.T. Kanpur
Submitted on 1.1.96

ABSTRACT

The primary objective of any transit system is to provide better level of service to its passengers. One of the good measures of level of service is the waiting time of passengers during their journey. The waiting time can be minimized through proper scheduling of the bus transit system. A good schedule minimizes both the overall transfer time (TT) of passengers transferring between different routes and the initial waiting time (IWT) of the passengers waiting to board a bus at their point of origin. In this study, a mathematical programming (MP) formulation of the scheduling problem for a transit system network is presented. The MP problem is large and nonlinear in terms of the decision variables which are real as well as integers. This makes it difficult for traditional methods to solve the problem. Hence, Genetic algorithms (GAs) which are powerful search and optimization methods, are used to solve the problem. Function based declarations and coding of variables in GAs allow an efficient reformulation of the original problem. The modified problem is computationally much simpler than the original problem. The results from a number of test problems show that the GAs are able to find optimal schedules with minimal computational resource.

ACKNOWLEDGEMENTS

Firstly, I endorse my deep feelings of gratitude to my adviser, Dr. Partha Chakroborty, Assistant professor for his invaluable guidence and immense help during all stages of my research. My sincere thanks are due to Dr. Kalyanmoy Deb, Assistant professor of Mechanical Engineering department for his useful suggestions.

The help rendered by Harikishan Reddy during my stay at IIT Kanpur is unforgettable. I thank Kaliprasad, Srinivas, Ravinder Reddy, Guptha, Vidyanath, Chakravorty, Sastry, Chender Rao for their help and good company. Finally, I thank all those who were indirectly part of my success.

B.Srinivas.

Dedicated to
my Parents

Contents

1 Introduction 1

2 Problem Statement 4

3 Traditional Formulation 8

3.1 Formulation 8

3.1.1 Assumptions 8

3.1.2 Decision variables 9

3.1.3 Objective Function 10

3.1.4 Constraints 14

3.1.5 Problems in Solving 16

4 Genetic Algorithms 18

4.1 Introduction 18

4.2 Differences with Traditional methods 19

4.3 Working Principle 21

4.3.1 Overview 21

4.3.2 Strings 22

4.3.3 Coding and Decoding 22

4.3.4 Evaluation 23

4.3.5 GA Operators 23

4.3.6 Termination Criteria 26

5 Revised Formulation **28**

5.1 Revised formulation 28

5.2 Implementation 30

6 Optimal Schedules **36**

6.1 GA Parameters 36

6.2 Network Parameters 37

6.3 Results 38

6.3.1 Arrival patterns 38

6.3.2 Distribution of buses 40

6.3.3 Optimal schedules 40

6.3.3.1 Optimal schedules for Network - I 41

6.3.3.2 Optimal schedules for Network - II 74

7 Conclusions **115**

References **117**

List of Figures

2.1 A Typical transit system network 5

2.2 Network - I 6

2.3 Network - II 7

3.1 Arrival pattern of passengers 13

4.1 A pseudo-code for a Simple Genetic Algorithm 21

5.1 Implementation of the Scheduling problem using GAs 31

6.1 Uniform arrival pattern 39

6.2 Non-uniform arrival pattern 40

6.3 Optimal schedule for TT consideration 43

6.4 Optimal schedule for IWT consideration (Uniform arrival pattern) 47

6.5 Optimal schedule for TWT consideration (Uniform arrival pattern) 50

6.6 Optimal schedule for IWT consideration (Non-uniform arrival pattern) 54

6.7 Optimal schedule for TWT consideration (Non-uniform arrival pattern) 57

6.8 Optimal schedule for TT consideration 61

6.9 Optimal schedule for IWT consideration (Uniform arrival pattern) 64

6.10 Optimal schedule for TWT consideration (Uniform arrival pattern) 67

6.11 Variance of Average total TT per w_{ijk}^{l} with w_{ijk}^{l}. 70

6.12 Optimal schedule for IWT consideration (Non-uniform arrival pattern) 71

6.13 Optimal schedule for TWT consideration (Non-uniform arrival pattern) 75

6.14 Schedule in one scheduling period 78

6.15 Schedule on either of a scheduling period 79

6.16 Optimal schedule for TT consideration 81

6.17 Optimal schedule for IWT consideration (Uniform arrival pattern) 85

6.18 Optimal schedule for TWT consideration (Uniform arrival pattern) 88

6.19 Optimal schedule for IWT consideration (Non-uniform arrival pattern) 92

6.20 Optimal schedule for TWT consideration (Non-uniform arrival pattern) 95

6.21 Optimal schedule for TT consideration 98

6.22 Optimal schedule for IWT consideration (Uniform arrival pattern) 102

6.23 Optimal schedule for TWT consideration (Uniform arrival pattern) 105

6.24 Optimal schedule for IWT consideration (Non-uniform arrival pattern) 108

6.25 Optimal schedule for TWT consideration (Non-uniform arrival pattern) 111

Notation

a^l_{ij} : Arrival time of the l^{th} bus of the j^{th} route at the i^{th} station.

d^m_{ik} : Departure time of the m^{th} bus of the k^{th} route at the i^{th} station.

dx^l_{ij} : Stopping time of l^{th} bus of the j^{th} route at the i^{th} station.

$E_{ij}(\tau)$: Locus of maximum $v^l_{ij}(t)$ for all l.

h_{ij} : Maximum headway of the j^{th} route at the i^{th} station, which is the bound on the difference in the arrival times of two successive buses.

L_{ij} : A factor controlling maximum height of the arrival pattern of passengers.

M : An arbitrary large number used in the formulation.

s^{min}_{ij} : Minimum stopping time of buses on j^{th} route at the i^{th} station.

s^{max}_{ij} : Maximum stopping time of buses on j^{th} route at the i^{th} station.

$t^l_{(i-1)ij}$: Travel time of the l^{th} bus of the j^{th} route between the $(i-1)^t h$ and i^{th} stations.

T : Maximum transfer time.

v^l_{ij} : Arrival pattern of passengers for the l^{th} bus on j^{th} route at the i^{th} station.

w^l_{ijk} : Transfer volume from the l^{th} bus of j^{th} route to the k^{th} route at the i^{th} station.

x^l_{ij} : Headway between l^{th} and $(l-1)^{th}$ bus on the j^{th} route at the i^{th} station.

α^l_j : Number of bits for the l^{th} headway on the j^{th} route.

δ^{lm}_{ijk} : A binary variable which is one if a transfer from the l^{th} bus j^{th} route to the m^{th} bus of k^{th} route at the i^{th} station is possible and optimal, Otherwise δ^{lm}_{ijk} is zero.

λ_j : Number of bits stopping time on the j^{th} route.

μ : A factor controlling the peak of the arrival pattern of passengers.

ν_j : Range for stopping time on the j^{th} route.

ρ^l_j : Range for the l^{th} headway on the j^{th} route.

Chapter 1

Introduction

The routing and scheduling of vehicles and their crew is an important area for transit system planners. An efficient routing and scheduling of transit systems and their crew can save large amount of resources which can now be utilized to provide still better services. At the same time it can provide the users of the transit system with a better level of service.

In general, the transit planning process can be done sequentially in four stages. They are network route design, setting time tables, scheduling of vehicles to trips and assignment of drivers. It is necessary to plan all the four components simultaneously for the transit system to be efficient. But this planning process is very difficult and complex and therefore require separate planning of each component, with the outcome of one fed as an input to the next component [3]. The present study deals with setting time tables, i.e. scheduling of bus transit systems.

Transportation contributes to the economic, industrial, social and cultural development of any country. It involves safe and efficient movement of people and goods. An efficient transit system provides less waiting time to its users. Scheduling of bus transit systems plays a major role in increasing the efficiency of the transit systems. A good schedule minimizes both the waiting time of passengers waiting to board a bus at their point of origin and the transfer time of passengers transferring between different routes at different stations in a transit network.

The attributes which make transfers inconvenient are the discomfort of boarding a new bus and the waiting for the arrival of that transfer vehicle. There are two ways to reduce waiting time for transfers in a transit network. The first one is the decrease of headway on certain routes. It involves an increase in the number of buses and drivers required to operate these routes and hence the operating costs increase. The second one is transfer coordination which includes *timed transfer* and *transfer optimization*. Timed transfer is a scheduling procedure whereby particular buses are scheduled to meet at certain transfer points, within given time windows called contact periods. Layover is added at the transfer points to ensure contact even if some of the buses are late. Due to this layover also the operating costs increase. "Timed transfer is inappropriate for a large transit network with decentralized transfers, where transfers occurs virtually at any intersection between two routes"[2]. In transfer optimization, buses are scheduled in order to minimize some objective function measuring the overall inconvenience of transfers in the network. In the present study, the objective function includes not only transfer time but also the time passengers spend in waiting before boarding a bus at their point of origin.

However, it is very difficult to develop such an optimal schedule even for a small transit network [13]. This difficulty is due to the large number of variables and constraints, discrete nature of variables and non-linearities involved in the objective function and the constraints. In this study, Genetic algorithms (GAs) are used to solve the scheduling problem. GAs are particularly suited for the scheduling problem as they allow a very efficient reformulation of the problem, which reduces the difficulties mentioned above. In an earlier work [16], Subrahmanyam showed the viability of using GAs as a solution tool. The thrust of the present study is to show the viability of using GAs to solve the network wide scheduling problem. Analyses for various cases are done and the results obtained are discussed. The results show the efficacy of GAs as the solution tool for the development of network wide optimal transit schedules.

This thesis is organized into seven chapters of which this is the first. Chapter 2 states the scheduling problem studied here. The traditional formulation of the scheduling problem and its description are presented in Chapter 3. Chapter 4 introduces Genetic algorithms, an optimization technique used to solve the scheduling problem. The modified formulation of the scheduling problem using GAs is presented in Chapter 5. The optimal schedules obtained for various cases and the discussion of the results are given in Chapter 6. The last chapter presents conclusions from this study.

Chapter 2

Problem Statement

In this chapter, the problem of optimal scheduling of transit systems is described.

On a typical transit network various routes operate. These points occasionally intersect at points referred to as *transfer points.* Figure 2.1 shows such a transit system network. The solid lines represent the various routes in the network, the dots on the lines represent the stops on the routes and the circled intersections represent the transfer stations in the network. The objective is to determine a schedule for the buses on each route which provides a better *Level of service (LOS)* to the passengers moving on the network within the available resources. One of the measures of the LOS is the waiting time, the amount of time a passenger spends in waiting during his journey from origin to destination. A lesser waiting time obviously translates to a better level of service.

At any transfer station on a transit network two types of passengers exist; for some the station is their point of origin and for others the station is a point where they wait to board a bus on a particular route after being dropped off by a bus on another route. Consider the Station A in the network shown in the Figure 2.1. Two kinds of passengers wait at this station. The first type of passengers have come to Station A and possibly wait to board a bus on any one of the routes at the station. The time a passenger spends in waiting at their point of origin is known as *initial waiting time*(IWT). The other type of passengers possibly come from Station D and go to any destination P, first come to Station A and transfer to a bus on route towards

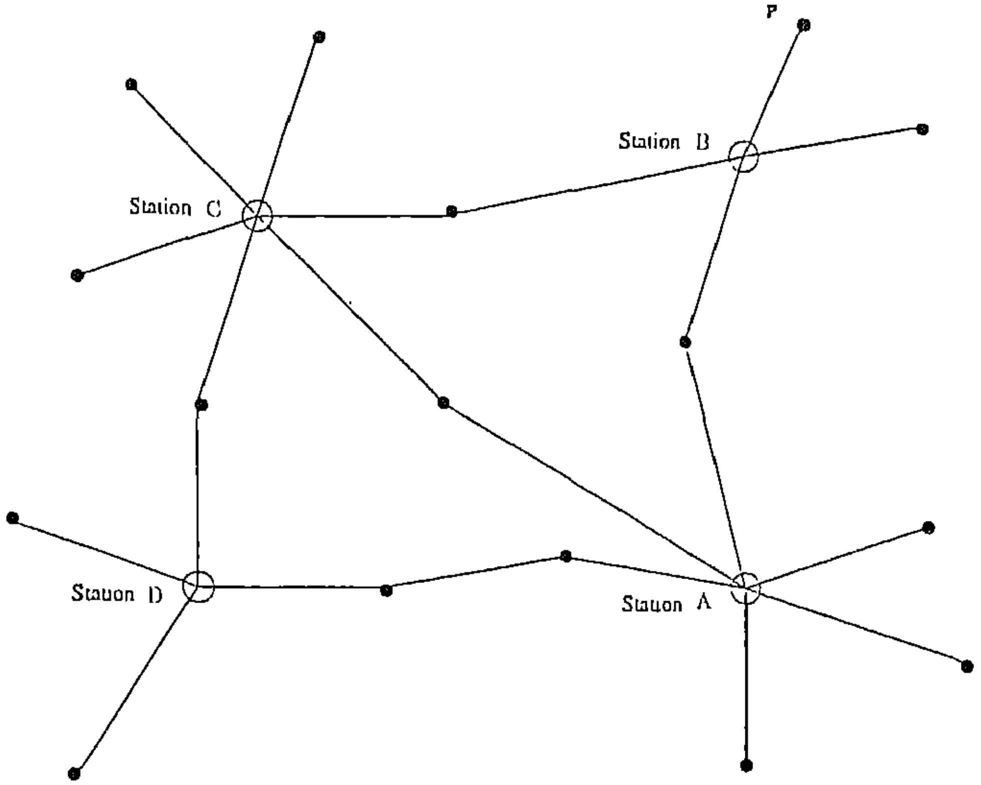


Figure 2.1: A Typical transit system network

P. The time a passenger spends in transferring from one route to another at different stations is known as *transfer time*(TT). The total time a passenger have to wait(TWT) is therefore the sum of the IWT and TT (where TT is zero if no transfer is involved). A good schedule therefore be one which will give a minimal value for the TWT.

It should be noted that the development of a good schedule(or any feasible schedule) proceeds under certain resource constraints as well as service related constraints. Though, one could possibly come up with various such constraints, the most important of these are considered here. The following enumerates the important resource constraints and service related constraints which are to be taken into account while developing a schedule.

Resource Constraints

1. Fleet size : Any transit system operator has a fixed fleet of buses. Thus the schedules should be developed based on the available fleet.
2. Capacity limitations : A bus may not accommodate all passengers

waiting to board the bus.

Service related Constraints

1. Minimum stopping time : A bus cannot start as soon as it arrives a station. It has to stop for a certain period of time.
2. Maximum stopping time : A bus cannot stop for more than a certain period of time at a station.
3. Maximum transfer time : A passenger cannot be allowed to wait for more than a certain period of time for the transfer.

In this thesis, two networks of increasing complexity are chosen to illustrate the efficacy of the proposed algorithm. The networks are briefly described here.

Network - I

The first network is shown in the Figure 2.2. There are three Stations 1, 2 and 3 in the network and two common routes (routes connecting any two

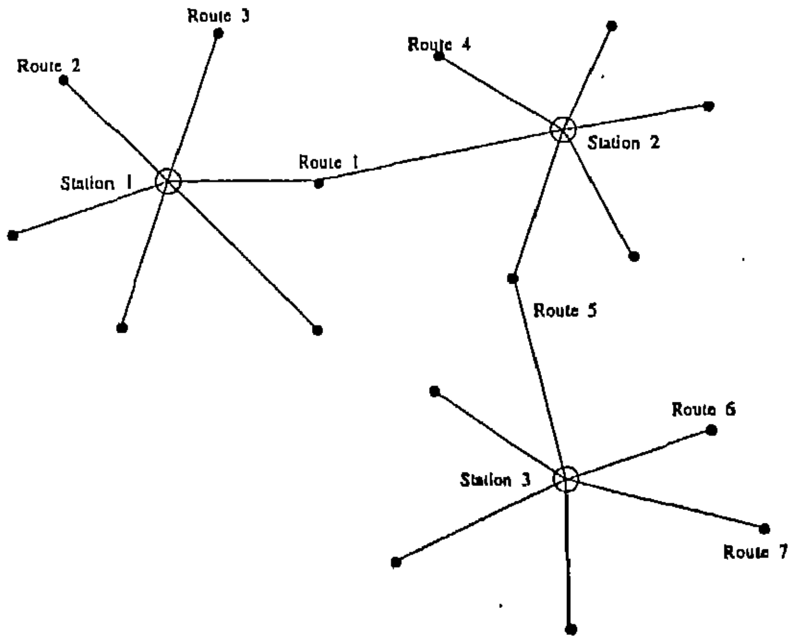


Figure 2.2: Network - I

stations) between Stations 1,2 and 2,3, remaining routes being independent with respect to other stations. As buses move on common routes from one station to another, the schedule developed for buses at one station obviously effect the schedule at another station. For example, the schedule at Station 1 will effect that at 2, which in turn will effect the schedule at Station 3. So the schedules should be developed for buses at all the stations simultaneously.

Network - II

The second network is shown in the Figure 2.3. This network is more complicated than Network-I as far as scheduling is considered as there is one more common route between Stations 1 and 3. The schedules at both Stations 1 and 2 will effect the schedule at 3. So the schedule at Station 3 should be developed based on schedules at Stations 1 and 2.

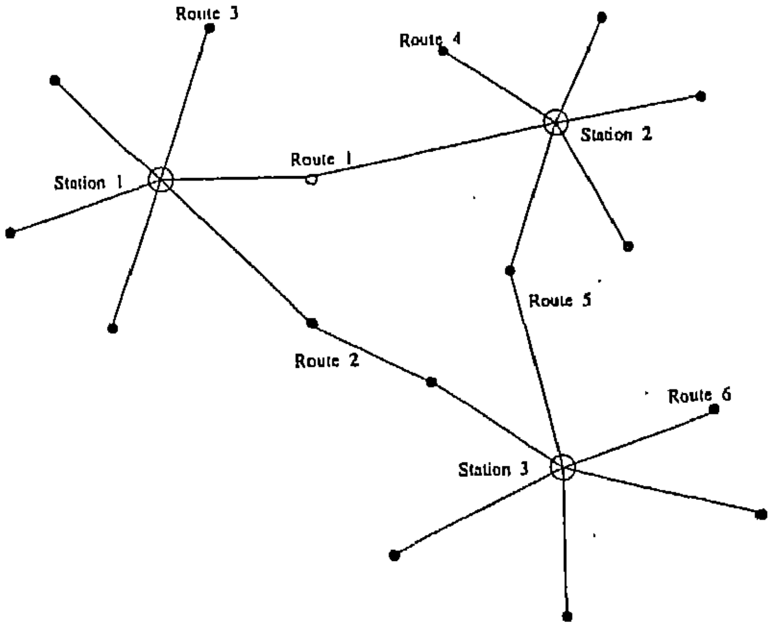


Figure 2.3: Network - II

In the next chapter, the traditional formulation of the above scheduling problem is presented.

Chapter 3

Traditional Formulation

In the previous chapter, the scheduling problem was described. In this chapter, the problem is formulated as a traditional mathematical program. First the formulation of the problem is described. Later, the difficulties in solving the problem with traditional methods are presented and the need for selecting an efficient optimization algorithm for solving the scheduling problem is stated.

3.1 Formulation

The formulation of the scheduling problem as a mathematical program is presented and discussed here. This section is divided into four subsections. The assumptions and decision variables of the problem are provided in the first two subsections. The third subsection describes the objective function while the last section explains the constraints.

3.1.1 Assumptions

In this section, the various assumptions made while formulating the scheduling problem are listed. The assumptions are :

1. The fleet size for all the routes at all the transfer stations are known.

2. Minimum stopping time : A bus has to stop for a period of atleast s_{ij}^{min} at the i^{th} station.
3. Maximum stopping time : A bus cannot stop for a period of more than s_{ij}^{max} at the i^{th} station.
4. Maximum transfer time : No passenger on the transit network should have to wait more than a time T at a transfer station.
5. Passengers are assumed to transfer from a bus of a route to the next available bus of the route to which they desire to transfer.
6. The number of transfers in between any pair of buses of different routes is known. This is however treated as a variable and different cases like no transfers at the transfer station, small number of transfers and large number of transfers at the station are taken into consideration.
7. Passengers whose point of origin is the station itself arrive at the station in a certain known pattern. Various arrival patterns of passengers are analyzed.
8. The capacity of a bus is large enough to accommodate all the passengers waiting to board the bus.
9. The scheduling time range is constant at all the stations in a network.
10. The stopping time for all the buses on a route at a station is same.
11. The travel time for all the buses moving on a connecting route between any two stations is same.
12. Intermediate stops on a connecting route between any two stations are not considered.
13. The routes are assumed to be one way.

3.1.2 Decision Variables

The decision variables in the formulation given below are the arrival times a_{ij}^{l}, the departure times d_{ij}^{l} and δ_{ijk}^{lm} values. The variable δ_{ijk}^{lm} is a zero-one integer variable. A value of zero means that the transfer from the l^{th} bus of the j^{th} route to the m^{th} bus of the k^{th} route at the i^{th} transfer station is not possible or not optimal. A value of one means otherwise. With these variable definitions, the scheduling problem explained in the last chapter can be formulated as follows.

Minimize

$$\sum_i \sum_j \sum_k \sum_l \sum_m \delta_{ijk}^{lm}(d_{ik}^m - a_{ij}^l)w_{ijk}^l + \sum_i \sum_j \sum_l \int_0^{d_{ij}^l - d_{ij}^{l-1}} v_{ij}^l(t)(d_{ij}^l - d_{ij}^{l-1} - t)dt \tag{3.1}$$

Subject to

$$\begin{array}{lcll} \text{G1} & : & (d_{ij}^l - a_{ij}^l) \leq s_{ij}^{max} & \forall i,j,l \\ \text{G2} & : & (d_{ij}^l - a_{ij}^l) \geq s_{ij}^{min} & \forall i,j,l \\ \text{G3} & : & (a_{ij}^l - a_{ij}^{l-1}) \leq h_{ij} & \forall i,j,l \\ \text{G4} & : & (d_{ik}^m - a_{ij}^l)\delta_{ijk}^{lm} \leq T & \forall i,j,k,l,m \\ \text{G5} & : & (d_{ik}^m - a_{ij}^l) + M(1 - \delta_{ijk}^{lm}) \geq 0 & \forall i,j,k,l,m \\ \text{G6} & : & \sum_m \delta_{ijk}^{lm} = 1 & \forall i,j,k,l,m \\ \text{G7} & : & (a_{ij}^l - d_{(i-1)j}^l) = t_{(i-1)ij}^l & \forall i \geq 2,j,l \end{array}$$

3.1.3 Objective Function

The primary objective of developing an optimal schedule for the bus transit systems in a network is the improvement of Level of Service offered by the transit system to the users. One of the measures of LOS is the waiting time of a passenger at a station in the network. By minimizing the sum of the waiting time of passengers coming from one route to transfer to another route at a transfer station and the initial waiting time of passengers who wait to board bus at their point of origin, we can obtain a better LOS. In this section the objective function used in the formulation is explained.

Given a particular schedule the objective function evaluate the total waiting time of passengers (i.e. sum of the total transfer time of passengers who transfer and the initial waiting time of passengers whose point of origin is the transfer station itself) of all the passengers at all the transfer stations in a network of routes.

Consider a typical transfer station such as Station 1 shown in the Network-I of Figure 2.2. There are three routes intersecting at the station. Two types of passengers wait at the transfer station. Some passengers have arrived at the station on a bus from one of the routes and wait to transfer to another route. These type of passengers are called *Transferring passengers.* For other passengers, the transfer station is their point of origin, and they wait to board a bus on one of the routes. These type of passengers are called *Non-transferring passengers.* Our objective is to arrive at a schedule such that the sum of the waiting times of the transferring passengers (Transfer time, TT) at the Stations 1, 2 and 3 and the waiting time of all the non-transferring passengers (Initial waiting time, IWT) at the Stations 1, 2 and 3 is minimal.

The transfer time of a passenger at the i^{th} transfer station depends on the departure time of the m^{th} bus on k^{th} route and the arrival time of l^{th} bus of j^{th} route at that station. The total transfer time is the sum of the all possible and optimal transfer times between any pair of buses on different routes at a station, considering all the stations in the network. It also depends on the number of passengers transferring from a bus on one route to that on another route i.e. w_{ijk}^{l}, on the number of stations in the network and the number of routes at each of the stations. The initial waiting time of passengers at a station depends on the arrival pattern of passengers for each route and the headway on the routes (which is the difference between the arrival times of successive buses on a route). The overall initial waiting time is the sum of the initial waiting times at each of the transfer stations in the network. So the objective function can be defined as the function of the variables $d_{ik}^{m}, a_{ij}^{l}, w_{ijk}^{l}, \delta_{ijk}^{lm}$ and $v_{ij}^{l}(t)$.

$$Waiting\ time\ function = f(d_{ik}^{m}, a_{ij}^{l}, w_{ijk}^{l}, \delta_{ijk}^{lm}, v_{ij}^{l}(t)) \tag{3.2}$$

In the scheduling problem, the objective function consists of two terms. The first term is the total transfer time (TT) for all the transferring passengers at all the stations where as the second term is the total initial waiting time (IWT) for all the non-transferring passengers at all the stations in the network.

Total Transfer time

At a transfer station in a network of routes, one type of passengers are those who want to transfer. These passengers came from their point of origin on one of the routes, get down at the station and wait for a bus on another route towards their destination. The total transfer time in the objective function is calculated as given below.

Consider a passenger who came on the l^{th} bus of the j^{th} route and wants to transfer to the k^{th} route at the i^{th} station. Let the arrival time of the l^{th} bus on j^{th} route be a^l_{ij}. A transfer to any bus on the k^{th} route can be possible if the bus departs after a^l_{ij}. But the passenger will board the bus whose departure time is closest to yet greater than a^l_{ij}. Let this be the m^{th} bus on the k^{th} route and its departure time from the station is d^m_{ik}. It therefore implies that the passenger has to wait for a time equal to $(d^m_{ik} - a^l_{ij})$. If δ^{lm}_{ijk} takes a value of one if $d^m_{ik} > a^l_{ij}$ and d^m_{ik} is the closest to a^l_{ij} amongst all the buses on the k^{th} route then the individual transfer time can can be written as :

$$Individual\ transfer\ time = (d^m_{ik} - a^l_{ij})\delta^{lm}_{ijk} \qquad \forall i,j,k,l,m \qquad (3.3)$$

In the above equation, we have assumed that one passenger is transferring from the l^{th} bus of the j^{th} route to the k^{th} route at the station i. But in reality, there may be more than one passenger who will transfer from the l^{th} bus of the j^{th} route to the k^{th} route at the station. If w^l_{ijk} is the number of passengers transferring from the l^{th} bus of the j^{th} route to the k^{th} route at the station, then the transfer time can be written as :

$$Transfer\ time = (d^m_{ik} - a^l_{ij})\delta^{lm}_{ijk}w^l_{ijk} \qquad \forall i,j,k,l,m \qquad (3.4)$$

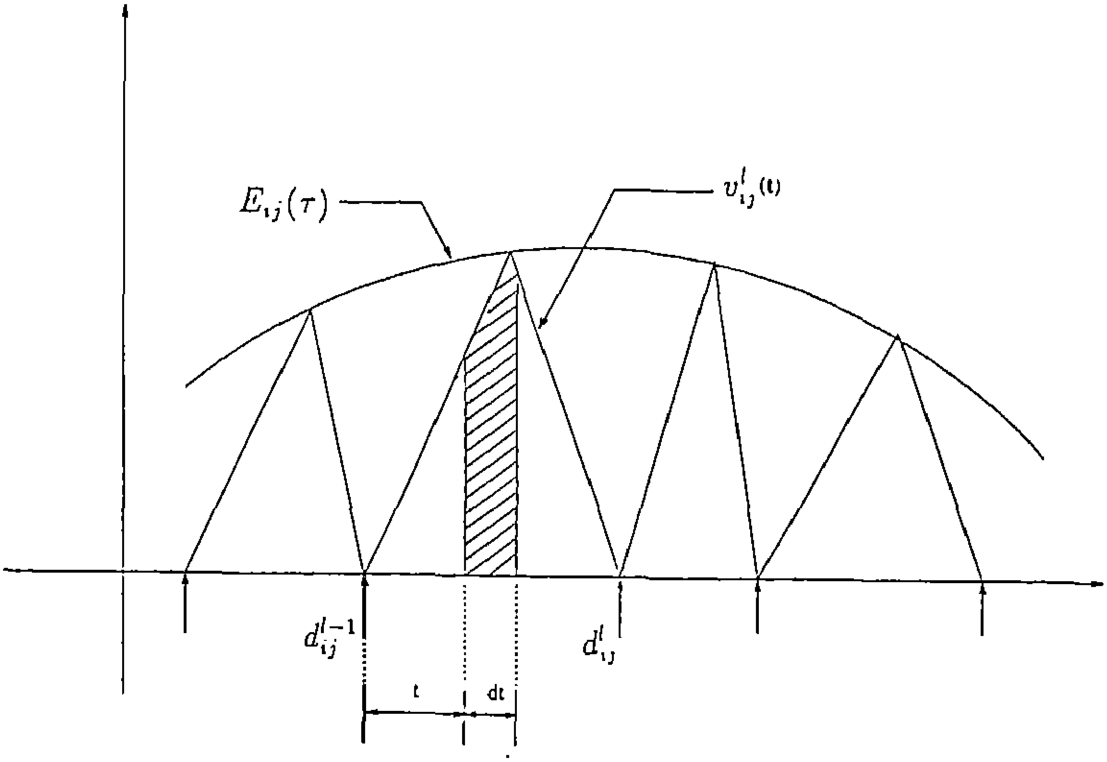
The total transfer time, TT can be obtained by summing the above term for all the buses on all the routes at all of the transfer stations in the network.

$$TT \ = \Sigma_i\Sigma_j\Sigma_k\Sigma_l\Sigma_m(d^m_{ik} - a^l_{ij})\delta^{lm}_{ijk}w^l_{ijk} \qquad \forall i,j,k,l,m \qquad (3.5)$$

Note that TT contains the product of two decision variables and hence it is non-linear.

Total Initial waiting time

The waiting time for the non-transferring passengers is referred to as the initial waiting time. Non-transferring passengers waiting to board the l^{th} bus of the j^{th} route at station i are assumed to have arrived at the station after the departure of $(l-1)^{th}$ bus of the same route. Thus the waiting time for a passenger arriving at time t $(d^{l-1}_{ij} \leq t \leq d^l_{ij})$ can be written as $(d^l_{ij} - d^{l-1}_{ij} - t)$. Further, if $v^l_{ij}(t)$ defines the arrival pattern of such passengers then the initial



Waiting time for passengers arriving in the shaded area $= v^l_{ij}(t)(d^l_{ij} - d^{l-1}_{ij} - t)dt.$

Figure 3.1: Arrival pattern of passengers

waiting time for all such passengers can be written as :

$$\textit{Initial waiting time} = \int_0^{d^l_{ij} - d^{l-1}_{ij}} v^l_{ij}(t)(d^l_{ij} - d^{l-1}_{ij} - t)dt. \qquad \forall i,j,l \quad (3.6)$$

The above discussion is further clarified with the help of Figure 3.1.

The total initial waiting time, IWT for all the passengers waiting on all the routes at each of the stations in a network is therefore given by

$$IWT = \sum_i \sum_j \sum_l \int_0^{d^l_{ij} - d^{l-1}_{ij}} v^l_{ij}(t)(d^l_{ij} - d^{l-1}_{ij} - t)dt. \qquad \forall i,j,l \quad (3.7)$$

It may be noted that the IWT term is also non-linear in the decision variables d^l_{ij}.

Figure 3.1 also illustrates the concept of the envelope function used here to describe the variations in demand over time. $E_{ij}(\tau)$ is such an envelope function. It is the locus of maximum $v^l_{ij}(t)$ for each route at a station and τ is the time at which $v^l_{ij}(t)$ attains a maximum and is related to the departure times as $\tau = d^{l-1}_{ij} + \mu(d^l_{ij} - d^{l-1}_{ij})$. Both Uniform and Non-uniform envelope functions were taken into consideration in this study. In the uniform arrival pattern of passengers the maximum $v_{ij}^l(t)$ is same for all buses on a route at a station. While in non-uniform pattern, the maximum $v_{ij}^l(t)$ is assumed to follow a β function given in the following equation.

$$E_{ij}(\tau) = \left(\frac{\tau}{\mu}\right)^{\mu}\left(\frac{S-\tau}{1-\mu}\right)^{1-\mu}\frac{L_{ij}}{S} \tag{3.8}$$

where L_{ij} is the scaling factor for the j^{th} route at the i^{th} station and is chosen so as to keep the number of passengers arriving at the station as per requirement, S is the maximum schedule time period, and the factor μ affects the location of the peak of the $v_{ij}^l(t)$.

3.1.4 Constraints

In the formulation of the scheduling problem given in the Equation 3.1 there are six constraints. The meanings of each of the constraints are briefly described in this subsection.

1. Maximum stopping time constraint :
   Constraint G1 : $(d_{ij}^l - a_{ij}^l) \leq s_{ij}^{max} \qquad \forall i, j, l$

   Stopping time at a station for a bus on any route should not be large enough so as to cause discomfort and frustration to the on-board passengers. Constraint G1 assures that the stopping time $(d_{ij}^l - a_{ij}^l)$ for the l^{th} bus on the route j at the i^{th} station should be less than or equal to the maximum stopping time, s_{ij}^{max}.

2. Minimum stopping time constraint :
   Constraint G2 : $(d_{ij}^l - a_{ij}^l) \geq s_{ij}^{min} \qquad \forall i, j, l$

   Stopping time at a station for a bus on any route should be such that passengers should be able to alight and board the bus. So the stopping time should not be too small. Constraint G2 assures that the stopping time $(d_{ij}^l - a_{ij}^l)$ for the l^{th} bus on the route j at the i^{th} station should be greater than or equal to the minimum stopping time, s_{ij}^{min}.

3. Headway constraint :
   Constraint G3 : $(a_{ij}^l - a_{ij}^{l-1}) \leq h_{ij} \qquad \forall i, j, l$

The time gap between any two successive buses on a route is known as headway. It should not be larger than a stipulated maximum headway in order to prevent large waiting times for passengers. By this constraint, we restrict the headway on a route to be less than or equal to the stipulated headway, h_{ij}.

4. Maximum transfer time constraint :
   Constraint G4 : $(d_{ik}^{m} - a_{ij}^{l})\delta_{ijk}^{lm} \leq T \qquad \forall i, j, k, l, m$

   It is inconvenient for a passenger to wait too long for a transfer. In general, a route with less travel demand will have less number of buses plying on it. Since the number of buses on all routes at a station may not be same, a passenger coming from a relatively high frequency route may have to wait for a long time to transfer to a low frequency route. As large transfer times would be highly inconvenient to the passengers, the maximum transfer time allowed in the schedule should be less than a certain stipulated maximum. Constraint G4 restricts the transfer time for any transfer to be less than or equal to the maximum transfer time T. As the constraint consists of the product of two decision variables, it is a non-linear constraint.

5. Transfer constraints :
   (a) Constraint G5 : $(d_{ik}^{m} - a_{ij}^{l}) + M(1 - \delta_{ijk}^{lm}) \geq 0 \qquad \forall i, j, k, l, m$

   This constraint assures that δ_{ijk}^{lm} is zero whenever a transfer from the l^{th} bus of j^{th} route to the m^{th} bus of the k^{th} route at the i^{th} station is not possible; i.e., whenever $d_{ik}^{m} < a_{ij}^{l}$. Under this condition $(d_{ik}^{m} - a_{ij}^{l})$ is negative, then $(1 - \delta_{ijk}^{lm})$ has to be one in order the constraint to be satisfied. i.e. δ_{ijk}^{lm} has to be zero. However, note that when $(d_{ik}^{m} - a_{ij}^{l})$ is positive, then according to this constraint $(1 - \delta_{ijk}^{lm})$ could be either one or zero.

   (b) Constraint G6 : $\sum_{m} \delta_{ijk}^{lm} = 1 \qquad \forall i, j, k, l, m$

   This constraint along with Constraint G5 assures that transfer from a particular bus of a particular route is made to only one of the possible buses of another route. Constraint G5 assured that for all infeasible transfers δ_{ijk}^{lm} will be zero. However, it did not assure that amongst the feasible transfers one and only one of the corresponding δ_{ijk}^{lm} should be one. Constraint G6 assures this. Note that this constraint does not assure that the δ_{ijk}^{lm} corresponding to the optimal transfer (for which $d_{ik}^{m} - a_{ij}^{l}$ is positive and smallest) be one. This is however, assured by the minimization of the objective function.

6. Travel time constraint :
   Constraints G7 : $(a_{ij}^{l} - d_{(i-1)j}^{l}) = t_{(i-1)ij}^{l} \qquad \forall i \geq 2, j, l$

   This constraint incorperates the dependency of the arrival time of a bus at a particular station on the departure time of the same bus from the previous station. This constraint assures that travel time for a bus on a common route is equal to the difference between the arrival time of the bus at a station and its departure time at the previous station.

3.1.5 Problems in Solving

The formulation given in Equation 3.1 of Section 3.1.2 is a non-linear mixed integer program. Further, even for a few routes and buses, the number of variables and constraints in the problem are very large. For example, consider a station where three routes are intersecting and each route has 10 buses during the scheduling period. In this case, there are 660 decision variables in the traditional formulation, out of which 600 are 0-1 integer variables. The total number of constraints are 1350 out of which 600 are non-linear constraints. With this many number of variables and constraints, the search space is very large and complex for a non-linear mixed integer problem. For bigger transit networks, the problem becomes all the more difficult since the complexity for this case increases polynomially with the number of routes and buses [16]. It is reported by Chakroborty et al. that the above formulation could not be solved within any reasonable period of time [4].

The non-linearity exits in the terms TT and IWT and the Constraint G4. Removal of non- linearity (i.e., Linearization) may make the problem tractable using the traditional linear mixed integer algorithms like Branch and Bound method and Cutting plane method. Linearizing polynomial functions is to express them first as functions of 0-1 variables and then to introduce new variables to take the place of product terms, simultaneously introducing auxiliary constraints to ensure that the new variables will assume the appropriate values. In general a non-linear problem cannot be linearized.

However certain types of non- linearity arising due to the product of two variables could be linearized. In the above formulation, the TT term and the non-linear constraint can be linearized using Glover technique [8]. However, the IWT term is not amenable to linearization. Even after neglecting the IWT term and linearizing the rest of the problem, the complexity remains large. The benefit obtained through linearization is offset by the increase in the number of variables and the number of constraints necessary for such linearization. It is difficult to solve the scheduling problem even in its simplified and linearized version by traditional methods mainly because of the large and complex search space [4].

The inadequacies of the traditional methods in solving the scheduling problem lead to the search for a better algorithm for solving such problems. Owing to the initial success of Chakroborty et al. in obtaining solutions to simple scheduling problems using *Genetic Algorithms* (GAs), the same is used here to explore its efficacy in solving larger and more realistic scheduling problems. The next chapter gives a brief introduction to Genetic algorithms and their advantages over traditional methods.

Chapter 4

Genetic Algorithms

In the previous chapter, the traditional formulation of the scheduling problem was described and the problems in solving the problem with traditional methods were presented. In this chapter, the Genetic algorithms are introduced first and the differences with traditional methods and their mechanism are given in the subsequent sections.

4.1 Introduction

Genetic algorithms (GAs) were developed by John Holland and his colleagues at the University of Michigan in 1965. GAs are adaptive search and optimization algorithms based on the mechanics of natural selection and natural genetics. They are very different from traditional search and optimization methods used in engineering design problems. Because of their simplicity, ease of operation, minimal requirement and global perspective; GAs have been successfully used in a wide variety of problems domains.

GAs combine Darwin's theory of survival of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm with some of the innovative flair of human search. In every generation (explained later), a new set of artificial creatures(strings) is created using bits and pieces of the fittest of the old; an occasional new part is tried for good measure. While randomized, Genetic algorithms are no simple random walk. They efficiently exploit historical information to speculate on new search points with expected improved performance. The GA is a search procedure that uses random choice as a tool to guide a highly exploitative search through a coding of parameter space.

Genetic algorithms are theoretically and empirically proven to provide robust search in complex spaces. Their robustness i.e. the balance between the efficiency and efficacy necessary for survival in many different environments makes them wide spread applications in business, scientific and engineering circles. These algorithms are computationally simple yet powerful in their search for improvement. Further, they are not fundamentally limited by restrictive assumptions about the search space (assumptions concerning continuity, existence of derivatives, unimodality and other matters).

4.2 Differences with Traditional Methods

Genetic algorithms are different from traditional optimization and search procedures in four ways :

1. GAs work with a coding of the parameter set, not the parameters themselves.
2. GAs search from a population of points, not from a single point.
3. GAs use payoff(objective function) information, not derivatives or other auxiliary knowledge.
4. GAs use probabilistic transition rules, not deterministic rules.

Genetic algorithms require the natural parameter set of the optimization problem to be coded as a finite length string over some finite alphabet. Since they manipulate decision or control variable representations at the string level to exploit similarities among high performance strings. Other methods usually deal with functions and their control variables directly. One of the advantages of working with a coding of variable space is that the coding discretizes the search space, even though the function may be continuous. A discrete or continuous function may be tackled using GAs, as the function values at various discrete points are required. Though there are many ways for coding of variables, Binary coding is mostly used.

Genetic algorithms work from a population of points; many other methods work from a single point. In many conventional optimization methods, a single point in the decision space is selected first and some transition rules are used to move to the next point. This point to point method is dangerous because it may locate a false peak i.e. local optimum in multimodal search spaces. But GAs start from a population of points simultaneously climbing many peaks in parallel, thus the probability of finding a false peak is reduced. Because there are many strings that are processed simultaneously and used to update any string in the population, it is very likely that the expected GA solution may be a global solution.

Many search techniques require auxiliary information to direct their search. For example, Gradient techniques require derivatives in order to climb the current peak. However GAs doesn't require any such kind of information. They remain general by exploiting information available in any search direction. To perform an effective search, they require only payoff values associated with individual strings.

The basic problem with most of the traditional methods is that there are fixed transition rules to move from one point to another. So these methods can only be applied to special class of problems where any point leads to the desired optimum. That is why these methods are not robust and simply can not be applied to a wide variety of problems. But the GAs use probabilistic rules to guide their search i.e. the transition rules of GAs are stochastic. However, a distinction exists between the randomized operators of GAs (explained in the next section) and other methods that are simple random walks. Genetic algorithms use random choice to guide a highly exploitative search towards regions of search space with likely improvement. This characteristic of GAs permits them to be applied to a wide class of problems giving them the robustness that is very useful in very sparse nature of engineering design problems.

Taken together, these four differences - direct use of coding, search from a population, blindness to auxiliary information and randomized operators-contribute to a Genetic algorithm's robustness and resulting advantage over other commonly used techniques. In the next section, the working principle of GAs is given.

4.3 Working Principle

In this section, an overview of GA mechanism is presented first. The next subsection explains about the strings used in GA. The coding and decoding of strings and the procedure for their evaluation are explained in third and fourth subsections respectively. The last two subsections describes the operators used in GAs to get new population of strings and GAs termination criteria respectively.

4.3.1 Overview

Figure 4.1 illustrates a pseudo-code for a genetic algorithm. GAs begin with a population of strings created randomly. Thereafter, each string in the population is evaluated. The population is then operated by three main operators - reproduction, crossover and mutation - to create a hopefully better population. The population is further evaluated and tested for termination. If the termination criteria are not met, the population is again operated by the above three operators and evaluated. This procedure is continued until the termination criteria are met. One cycle of these operators and the evaluation procedure is known as a *generation* in GA terminology.

```
Begin
    Initialize population;
    Evaluate population;
    repeat
        Reproduction;
        Crossover;
        Mutation;
    until (termination criteria);
End.
```

Figure 4.1: A pseudo-code for a Simple Genetic Algorithm

In the following pages, all the above components in the pseudo-code are described after discussing the strings in GAs and their coding and decoding.

4.3.2 Strings

Genetic algorithm starts with initialization of strings. These strings of artificial genetic systems are analogous to chromosomes in biological systems. The chromosomes are composed of genes which take some number of values called alleles. Similar to the genes in the natural chromosomes which influence the physical properties of an individual, the values of 1 or 0 in each position of a binary string influence the coordinates of a point in search space as each string represents all the problem variables in the optimization problem considered.

The strings in the initial population of GA can be created by tossing of an unbiased coin. The successive coin flips(head=1, tail=0) can be used to decide genes in a string. This procedure is continued till all the strings are generated in the initial random population. The String length for a string i.e. the total number of bits (genes) is equal to the sum of the bits assumed for each of the problem variables in the optimization problem. The number of bits for a variable depends on the precision required (explained later).

4.3.3 Coding and Decoding

GAs work with coding of variables instead of variables themselves. One successfully used method of coding multiparameter optimization problems of real parameters is the concatenated, multiparameter, mapped fixed-point coding. In multivariable optimization, all the problem variables are coded in a single string and each variable will have a specific location in the string. Binary coding is most commonly used in GA applications. The precision of problem variables depends in the number of bits assumed for them in the string and the actual range in which they vary. The coding aspect is explained in the next chapter.

A string represents a point in the space and the decoded values of the string's contents will represent the coordinates. For example, if an integer variable varies from 10 to 41, it can be completely represented by a five digit string which can give 32 unique strings. The decoded value for string 00000 is 0, where as the decoded value for 11111 is 31, remaining all strings will have decoded values in between 0 and 31. Consider a string 10100. The decoded value for this string is $(0 \times 2^0 + 0 \times 2^1 + 1 \times 2^2 + 0 \times 2^3 + 1 \times 2^4)$ i.e. 20.

4.3.4 Evaluation

The decoded values from a string varies 0 to 2^l where l is the number of bits in the string. However, the actual values may not vary in that range. Then, the decoded values obtained from the strings should be mapped in the range in which the actual values will vary. Once all the variables are decoded they should be mapped as given below. Let ll and ul be the actual lower and upper bounds in which a variable vary and dc is the decoded value of the string, Then the mapped value is given by the following equation.

$$Mapped\ value = ll + \frac{(ul - ll) \times dc}{2^l - 1} \tag{4.1}$$

The precision of the variable is $\frac{(ul-ll)}{2^l-1}$

After all the variables are mapped, they can be used to calculate the objective function value (also called fitness value). In this way, all the fitness values for all the strings in a particular generation are calculated. Then the termination criteria is checked. If the termination criteria is not reached, the GA operators are applied on the present population to get a new set of population. The GA operators are described in the next subsection.

4.3.5 GA Operators

Genetic algorithms operate on population of strings, with the string coded to represent some underlying parameter set. A simple genetic algorithm that yields good results in many practical problems is composed of three operators - *Reproduction, Crossover and Mutation*, which are applied to successive string populations to create new string populations. These operators are very simple involving nothing more than random number generation, string copying and partial string exchanging. Despite their simplicity, the resulting search performance is wide ranging and impressive. These three operators are described below.

Reproduction

Reproduction is the first operator applied on a population. It is a process in which individual strings are copied according to their objective function values. Copying strings according to their fitness values means that the strings with a better value will have a higher probability of contributing one or more strings in the next generation. The operator is an artificial version of natural selection, a Darwinian survival of the fittest among string structures. The best string in the present generation will get more copies , the average stay even and the worst will die off in the next generation. It selects good strings in the population and forms a mating pool. There exists a number of reproduction operators in GA literature, but the essential idea is that above-average strings are picked from the population and duplicates of them are inserted in the mating pool. The commonly used reproduction operators are the *Proportionate Selection* operator and the *Tournament Selection.* The Tournament Selection operator is used in our study and is described below.

Tournament Selection

In Tournament Selection, some strings are picked at random and the better string is copied to the mating pool. This process is repeated till the required number of individuals equal to the population size are selected. The number of individuals considered for the selection of an individual is known as *Tournament size.* A tournament size of 2 is used generally in many applications and the same is used in our study. In this selection, two individuals are chosen at random in the current population and the better of them is selected with fixed probability ranging from 0.5 to 1.0. The strings thus selected by reproduction are kept in mating pool for applying the next operator, Crossover. Tournament selection can be used for both minimization and maximization problems without any transformation of the objective function.

Crossover

Crossover operator is applied next to the reproduction to the strings in the mating pool. After reproduction, the population is enriched with good strings from the previous generation, but does not have any new string. Crossover operator is applied to the population hopefully to create better strings. There exists a number of crossover operators like single point, multipoint crossover , uniform crossover. In all these operators, two strings are picked from the mating pool at random and some portion of strings are exchanged between the strings. In the single point crossover, this is performed by chosing a random site along the string and by exchanging all bits to the right of the crossing site as shown below.

$$\begin{matrix} 101|11 \\ 111|00 \end{matrix} \Longrightarrow \begin{matrix} 101|00 \\ 111|11 \end{matrix}$$

The crossover operator searches the parametric space by exchanging the information between two strings. The strings selected for applying crossover operator are called the parent strings and the strings obtained after crossover are known as child strings. In order to preserve some of the good strings found previously, crossover is usually performed on only some of the strings in the mating pool. The total number of strings to be participated in crossover is controlled by the *Crossover probability*, which is the ratio of total number of strings selected for mating and the population size. The crossover probability is generally is high, which varies from 0.90 to 0.95 in many applications.

If appropriate site is selected good child strings are obtained. However, as this site is not known beforehand, a random site is selected. With random site, the child strings may or may not be good. If good strings are obtained, there won't be any problem. Otherwise, the bad strings created will not survive beyond next generation due to the application of reproduction operator in the next generation. Crossover operator is mainly responsible for the search aspect of genetic algorithm.

Mutation

Mutation plays a secondary role in the operation of Genetic algorithm. It is needed because, even though reproduction and crossover efficiently exploit the search space, sometimes they lose some potentially useful genetic material (1's or 0's at particular locations). Mutation operator changes a 1 to 0 and vice versa with a small *Mutation probability.* The need for mutation is to keep diversity in the population. It is done bit-by-bit basis. Since this operator disrupts a string, the mutation probability is kept very low.

Mutation can be best explained with the help of the following example. Consider a population of size four. Let the strings are

1110
1010
1100
0100

Let the optimal string be 1111. By reproduction and crossover we will never be able to get the optimal string. But mutation may bring 1 in the last position of 1110 resulting in 1111. Then finally the problem will converge to the string 1111.

After applying the GA operators, a new set of population is created. Then, they are decoded and objective function values are calculated. This completes one generation of GA. Such iterations are continued till the termination criteria are reached.

4.3.6 Termination Criteria

When the average fitness of all the strings in a population is nearly equal to the best fitness, then the population is said to be converged. When the population is converged, the GA is terminated. The same can be done by fixing maximum number of generations, the number of generations at which population will converge. In Genetic algorithms, maximum number of generations is generally used as the termination criteria. The same is used in the present study.

In this chapter, a brief introduction to Genetic algorithms was given. The revised formulation of the scheduling problem with GAs and how GAs will reduce the complexity in the problem are presented in the next chapter.

Chapter 5

Revised Formulation

In the Chapter 3, the traditional formulation of the scheduling problem was presented and the Genetic algorithms were introduced in the previous chapter. In this chapter, the revised formulation of the problem is presented. It should be understood that such formulation is possible only due to the use of GAs as solution technique. While modified formulation is present in the first section, the second section deals with its implementation.

5.1 Revised Formulation

In this section, the original problem presented in the Equation 3.1 is reformulated. Before discussing the revised formulation, it is imperative that one understands the constraints in the original formulation. The constraints in the formulation given in the Equation 3.1 come under one of the three types given below.

1. Type I : Variable bounds.
2. Type II : Maximum transfer time.
3. Type III : Determination of δ_{ijk}^{lm}.
4. Type IV : Arrival time dependency.

Constraints G1, G2 and G3 come under the category of Type I. G1 and G2 provide the upper and lower bounds on stopping time on a route where as G3 provides an upper bound on the headway between any two consecutive buses on a route. Constraint G4 comes under the category of Type II and provides an upper bound on transfer time. The constraints G5 and G6 come under the category of Type III while the last constraint comes under Type IV. In the following the revised formulation is presented keeping in view this discussion of the constraints.

The revised formulation (GA formulation) uses a different set of variables than the original formulation. The variables used here are x_{ij}^l and dx_{ij}^l where

$$x_{ij}^l = (a_{ij}^l - a_{ij}^{l-1})$$
$$\text{and } dx_{ij}^l = d_{ij}^l - a_{ij}^l.$$

The variable definitions are used to effectively benefit from some of the features of GAs. The benefit of using these variable definitions become clear later. It is to be noted that the variable definitions used here cannot be used beneficially in the original formulation since one needs direct knowledge of arrival and departure times in order to calculate δ_{ijk}^{lm} as well as transfer times.

With the new variable definitions and the coding of variables possible in GAs, the constraints G1, G2 and G3 of the original formulation can be eliminated. One can now define strings for variables dx_{ij}^l which will map into the ranges $[s_{ij}^{min}, s_{ij}^{max}]$; and another set of strings for the variables x_{ij}^l which will map into the range $[\underline{x}_{ij}^l, h_{ij}]$ where $\underline{x}_{ij}^l$ is an appropriate lower limit for the headways, for example 0.

Constraints G5 and G6 considerably added to the complexity of the original formulation; because they use integer variables and Constraints G6 are equalities. However, the purpose of these constraints is straight forward. For a given set of values of i,j,k,l and m; they set the $\delta_{ijk}^{lm} = 1$ if $(d_{ik}^m - a_{ij}^l)$ is positive and minimum. Otherwise they set $\delta_{ijk}^{lm} = 0$. The above logic can be incorporated easily in Genetic algorithms since GAs allow the use of external procedures during the optimization process. By avoiding these Type III constraints, the complexity of the problem is substantially reduced. Constraint G7 can also be eliminated using external procedures. Thus, the only constraints that are left out are the Type II constraints. The scheduling problem now becomes the minimization of the objective function subject to the constraint G4 only. These Type II constraints are tackled using a bracket operator exterior penalty term [15].

In the following section, the implementation of the above revised formulation is discussed.

5.2 Implementation

In this section, the implementation of the scheduling problem using Genetic algorithms is presented. In order to aid in the understanding of the implementation a flow diagram of the process is shown in the Figure 5.1. It may be noted that the flow diagram does not include all the intricate details. Only those which illustrate the logic and aid in the understanding of the implementation have been included. The following discussion which is divided into six steps. is based on Figure 5.1.

Step I

In the first step, GA creates randomly a population of strings. Each string represents a schedule and substrings represent all the problem variables (the headways and the stopping times on all the routes at all the stations present in the network). If there are m_j buses on j^{th} route, there will be $m_j - 1$ headways and m_j variables for stopping times. Since the stopping times for all the buses on a route at a particular station are considered to be equal, there is only one variable for the stopping time for a route at a station. Thus there are m_j number of variables (the same number of substrings) for each route at a station. However, for a route common to two stations, the stopping times at each of the stations are assumed to be different. So if a route j is common to s stations, then for route j there are $m_j + (s - 1)$ variables. Hence for a network containing a total of N routes out of which n are common routes(routes common to atleast two stations) the string will contain $\Sigma_{j=1}^{N-n} m_j + \Sigma_{j=1}^{n}(m_j + s - 1)$ variables (or substrings). In a network where for all common routes s=2, there are a total of $\Sigma_{j=1}^{N} m_j + n$ variables or substrings.

The problem variables x_{ij}^{l}(headways) and dx_{ij}^{l}(stopping times) take only discrete values at a specified precision level. Any desired precision can be set in genetic algorithms. Higher the precision level, larger should be the string length. As binary coding is used to code the problem, the desired precision

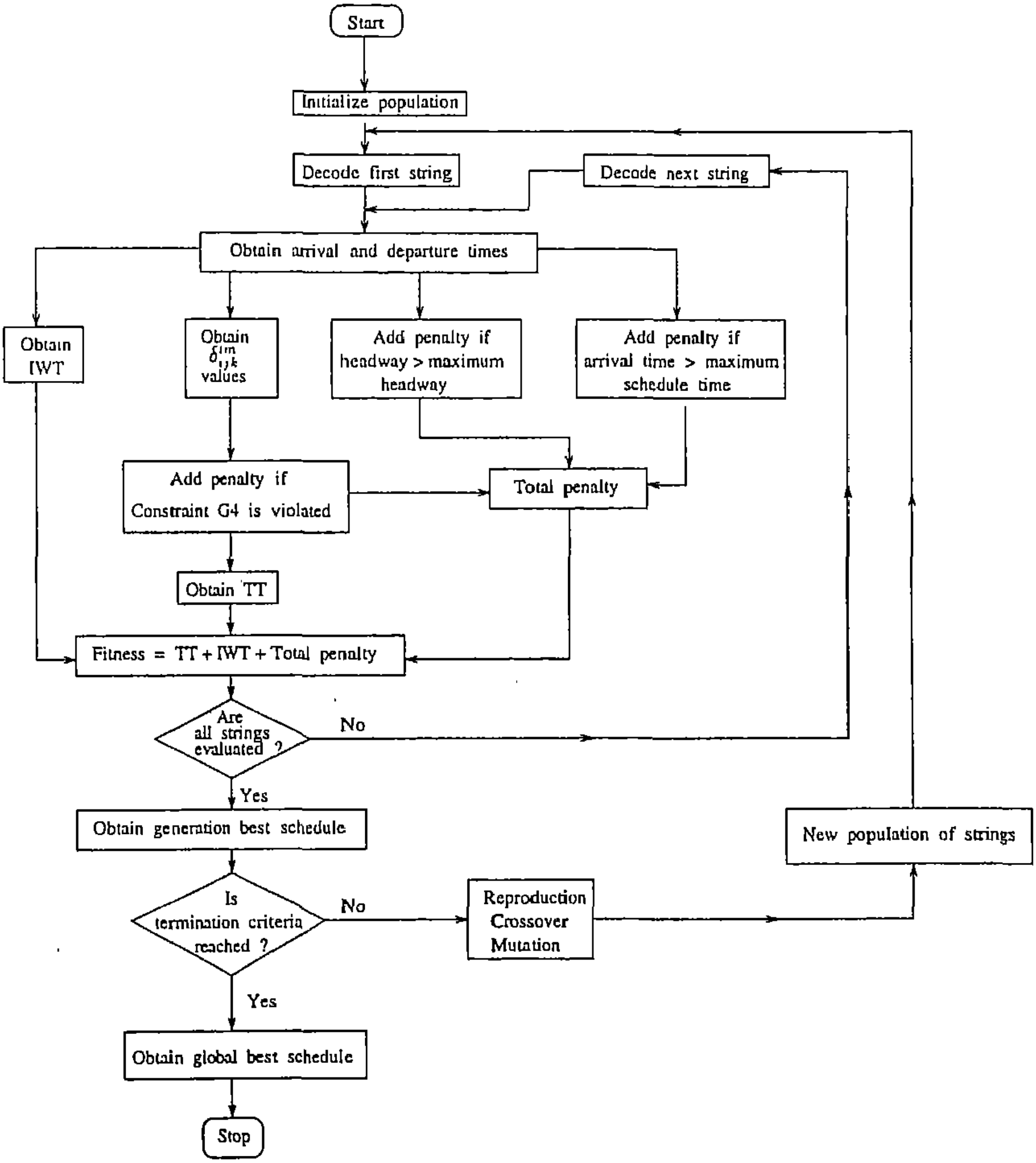

Start
Initialize population
Decode first string
Decode next string
Obtain arrival and departure times
Obtain IWT
Obtain δ_{ijk}^{lm} values
Add penalty if headway > maximum headway
Add penalty if arrival time > maximum schedule time
Add penalty if Constraint G4 is violated
Total penalty
Obtain TT
Fitness = TT + IWT + Total penalty
Are all strings evaluated ?
No
Yes
Obtain generation best schedule
Is termination criteria reached ?
No
Reproduction Crossover Mutation
New population of strings
Yes
Obtain global best schedule
Stop

Figure 5.1: Implementation of the Scheduling problem using GAs

can be set beforehand. Since all time tables which show the schedule of bus transit systems are generally specified upto one minute precision level, the same precision is used in the present study for various variables.

Since finite length strings are used to code the schedule, the range in which the decision variables vary should be fixed. The ranges may be different for different variables. If the range for the headway x_{ij}^l is ρ_j^l, then the number of bits(α_j^l) required to code the variable is

$$\alpha_j^l = \log_2 \rho_j^l \tag{5.1}$$

The unit interval of ρ_j^l should be so chosen that the α_j^l is an integer. For example, if the range for headway is 16 minutes and the unit of ρ_j^l is chosen to be $\frac{16}{2^{\alpha_j^l}}$, then the substring required to code the headway is α_j^l. Thus if the unit of ρ_j^l is 1 the substring length is 4; if the unit of ρ_j^l is 2 the substring length is 3; and so on. In this problem, the maximum headway was taken as 45 minutes and the minimum headway was taken as 14 minutes. For 1 minute precision level and for the range of 32, the substring length is 5. However, the first headway was assumed to vary from a minimum value of 0 to a maximum of 31 minutes(since no bus precedes the first bus). For this headway also, the substring length is 5. And the stopping time was assumed to be vary between 2 and 5 minutes. For this 4 minutes range (ν_j)with a precision level of 1 minute, the number of bits for stopping time (λ_j) is 2.

Thus, based on the the total number of variables in the optimization problem and the bits for each of the variables, the overall string length (L) can be calculated. The overall string length (L) which is the summation of all the substring lengths is given by $\sum_{j=1}^{N} \sum_{l=1}^{m_j-1} \alpha_j^l + \Sigma_{j=1}^{N+n} \lambda_j$. In the population all the strings will have the number of bits equal to L.

Step II

In the second step, each string or more correctly, each of the substrings of a string are decoded and the actual parameters are obtained. The decoding is done by mapping the binary substring to its binary decimal (though discrete) counterpart. For example, consider a two bit substring (representing the stopping time) 11. The decoded value for the stopping time is 5 ($1 \times 2^0 + 1 \times 2^1 +$ lower limit of stopping time).

Step III

The actual parameter values obtained in the Step II are used to calculate the arrival times (a_{ij}^l) and the departure times (d_{ij}^l) for all the buses on all the routes at each of the stations. The δ_{ijk}^{lm} values for all i, j, k, l and m are obtained using the values a_{ij}^l and d_{ij}^l. If p be the number of stations in a network, q_i be the number of routes at the i^{th} station and r_{ij} be the number of buses on j^{th} route at the i^{th} station, the computations are done with the help of the following loop.

```
Begin
for all stations (i=1 to p)
        for all routes (j=1 to $q_i$) compute
        $a_{ij}^0 = 0$
                for all buses (l=1 to $r_{ij}$) compute
                $a_{ij}^l = a_{ij}^{l-1} + x_{ij}^l$
                $d_{ij}^l = a_{ij}^l + dx_{ij}^l$
        for all combinations of routes (j,k) and buses (l,m)
        $\delta_{ijk}^{lm} = 0.$
        for all combinations of routes (j,k) and buses (l,m)
        repeat
                if($d_{ik}^m \leq a_{ij}^l$) then $\delta_{ijk}^{lm} = 0$
                else $\delta_{ijk}^{lm} = 1.$
        until( $\delta_{ijk}^{lm} = 1$)
End.
```

In this study, a finite length scheduling period of P is maintained at each of the stations in a network. The last bus on each route must arrive at $S_{ij} + P$ where S_{ij} is the starting time of the schedule on the j^{th} route at the i^{th} station. So, the last bus arrival is fixed at $S_{ij} + P$ and hence the last headway is not a decision variable. However, it might so happen that the last headway computed as the difference of $S_{ij} + P$ and the departure time of penultimate bus is greater than h_{ij} (thereby making the solution unacceptable). To eliminate such infeasibilities in the optimal solution, a penalty term with penalty parameter 10^7 is added if the last headway computed is greater than h_{ij}. Also any solution with arrival time of a bus greater than $S_{ij} + P$ is considered as infeasible. To avoid such infeasibilities, a penalty term with penalty parameter 10^7 is added to the objective function value, whenever such infeasibilities in the optimal solution occur. Also in this step, the IWT term in the objective function is calculated using the d^l_{ij} values for a particular arrival pattern of passengers.

Step IV

From the δ^{lm}_{ijk} values and the d^m_{ik} values obtained in Step III, the transfer time for each bus can be easily computed. If the transfer time, $(d^m_{ik} - a^l_{ij}) > T$(in this study, T = 30 minutes), then a bracket operator exterior penalty term is added to the objective function [15]. In the exterior penalty function method a penalty term $a\Sigma_{i,j,k,l,m}g(x)^b$ is used, where a is a penalty parameter, the exponent b is non-negative constant. The function g(x) is equal to $(d^m_{ik} - a^l_{ij})\delta^{lm}_{ijk} - T$ if $(d^m_{ik} - a^l_{ij})\delta^{lm}_{ijk} > T$, otherwise it is zero. The value of $a\Sigma_{i,j,k,l,m}g(x)$ increases as the b^{th} power of the amount by which the constraint is violated. As the present scheduling problem is a minimization problem, i.e. minimization of the sum of the objective function value and the total penalty, a positive value is used for b. This ensures that good solutions will not violate the constraint. A value of 2 was taken for b and 10^7 for a while minimizing the objective function.

Step V

In this step, the objective function value is evaluated for each string in the population. The minimum and average fitness values in the generation are calculated after evaluating all the strings. The schedule corresponding to the minimum fitness value is considered as the best schedule in that generation. The global best schedule upto the generation in question is found by comparing the minimum fitness values in all the previous generations. If the current generation is equal to the maximum number of generations assumed, the program is terminated and the global best schedule is considered as the optimal solution. A population may also converge to an optimal solution before the maximum number of generations. If this happens then also the process is terminated. By convergence it is meant that the average fitness of a generation be quite close to the best in that generation.

Step VI

If the process is not deemed to be converged or the number of generation is less than the maximum number of generations, the GA cycle is continued. The GA operators reproduction, crossover and mutation are applied on the present population to obtain a new set of strings. The above procedure is continued until the termination criterion is met.

In this chapter, it was seen that how efficiently GA formulation reduces the complexity in the scheduling problem. It may be reiterated that procedure based function declarations and coding of problem variables of GAs helped in reducing the complexity of the scheduling problem. The optimal schedules obtained using the revised formulation given in this chapter for various cases are presented and discussed in the next chapter.

Chapter 6

Optimal Schedules

In this chapter, the efficacy of the Genetic algorithms in solving the scheduling problem presented is shown. while developing schedules, various cases are considered for the two networks presented earlier (Figure 2.2 and Figure 2.3). In the following sections, the optimal schedules obtained for the cases for the two networks are presented. These schedules are best schedules obtained in 1000 generations of genetic algorithms. The various parameters used for the genetic algorithms and the parameters selected for the networks while developing the schedules are given in the first two sections. The results for the two networks are presented in the last section.

6.1 GA Parameters

The parameters for the genetic algorithms are itemized as follows.

- Simple Genetic Algorithm (SGA) is used.
- Number of bits used for coding headways = 5.
- Number of bits used for coding stopping times = 2.
- String length varies with the number of variables present.
    1. String length for Network I = 333, with equal resource on each route (defined later).

2. String length for Network I = 323, with unequal resource on each route (defined later).
3. String length for Network II = 288, with equal resource on each route.
4. String length for Network II = 288, with unequal resource on each route.

- Maximum population size = 1000, population size should be of the order of the chromosome length.
- Maximum number of generations = 1000.
- Cross over probability = 0.95; cross over probability is generally between 0.90 - 0.95. Single point crossover is used.
- Mutation Probability = 0.002; mutation probability is generally between 0.1/string length to 1.0/string length. Bitwise mutation is used.
- Binary tournament selection is used.

6.2 Network Parameters

In the following, parameters related to the networks studied here and the various bounds used on the variables pertinent to the scheduling are itemized.

- Number of transfer stations = 3.
- Number of routes at each of the stations = 3.
- Maximum schedule time window at each station = 240 minutes.
- Number of buses considered on each route = 8, 10 or 12.
- Travel times assumed :
    1. Between Stations 1 and 2 = 30 minutes.
    2. Between Stations 1 and 3 = 50 minutes.
    3. Between Stations 2 and 3 = 40 minutes.

- Maximum transfer time = 30 minutes.
- Minimum and Maximum Stopping times = 2 and 5 minutes.
- Headway Range :
  1. For first bus on each route = 0 - 31 minutes.
  2. For all remaining buses = 14 - 45 minutes.
- $\mu = 0.75$.
- $L_{ij} = 1.6$, 2.0 and 2.4 for routes with 8, 10 and 12 buses respectively.
- Various different values for w_{ijk}^{l} are used.

6.3 Results

In this section, the optimal schedules obtained for the two networks shown in Figure 2.2 and Figure 2.3 are presented. The schedules are derived for two different arrival patterns of passengers and two different distributions of buses on the routes. In the following, the different arrival patterns and distribution of buses used in this study are described.

6.3.1 Arrival Patterns

The two arrival patterns used are termed as *uniform arrival pattern* and *non-uniform arrival pattern.*

Uniform arrival pattern

Figure 6.1 shows an uniform arrival pattern. The arrows in the figure are the departure times of the buses. As can be seen from the figure, the pattern of arrival of passengers is triangular. The pattern is termed uniform because the height of the triangles remain the same over the entire scheduling period.

The area of each of the triangle is equal to the number of passengers coming in that period and the sum of areas of all the triangles is equal to the total number of passengers arriving on that route during the scheduling

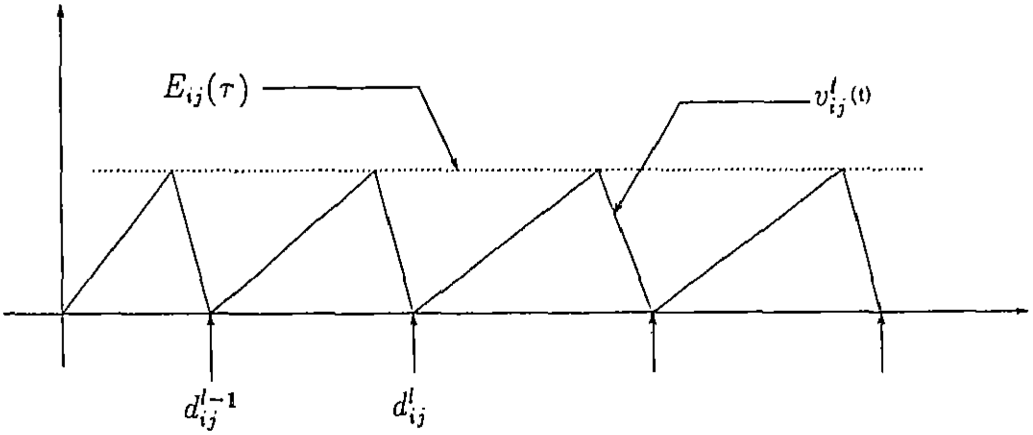


Figure 6.1: Uniform arrival pattern

period. The moment of area of a triangle about the departure time of a bus for which passengers wait gives the total waiting time for all the passengers. And the sum of the moments of areas will give the total initial waiting time for all the passengers on that route during the entire scheduling period.

Non-uniform arrival pattern

Figure 6.2 shows a non-uniform arrival pattern. As opposed to the uniform arrival pattern, in this the heights of the triangles vary over the entire scheduling period. As given in Chapter 3, the locus of the peaks ($E_{ij}(\tau)$) is a β function; this is also shown in the Figure 6.2.

For the examples solved in this thesis, the peak of the β function without any loss in generality, is assumed at $\frac{3}{4}^{th}$ (i.e. $\mu = 0.75$) of the scheduling period i.e. more number of passengers arrive in the latter half of the scheduling period. The maximum height of the β function is assmed to be equal to 2.0 (i.e.$L_{ij} = 2.0$) when the number of buses on each route are equal. For the case in which there are different number of buses on different routes, the value of L_{ij} for a route is fixed proportional to the number of buses on the route. As in the case of uniform arrival pattern, the sum of the areas of triangles gives the total number of passengers on that route where as the the sum of the moments of areas gives the waiting time for all passengers on that route.

The heights of the triangles in the earlier case is kept at 1.4417 so that the total number of passengers on a particular route is equal to the number of passengers on the same route under non-uniform arrival pattern. This is

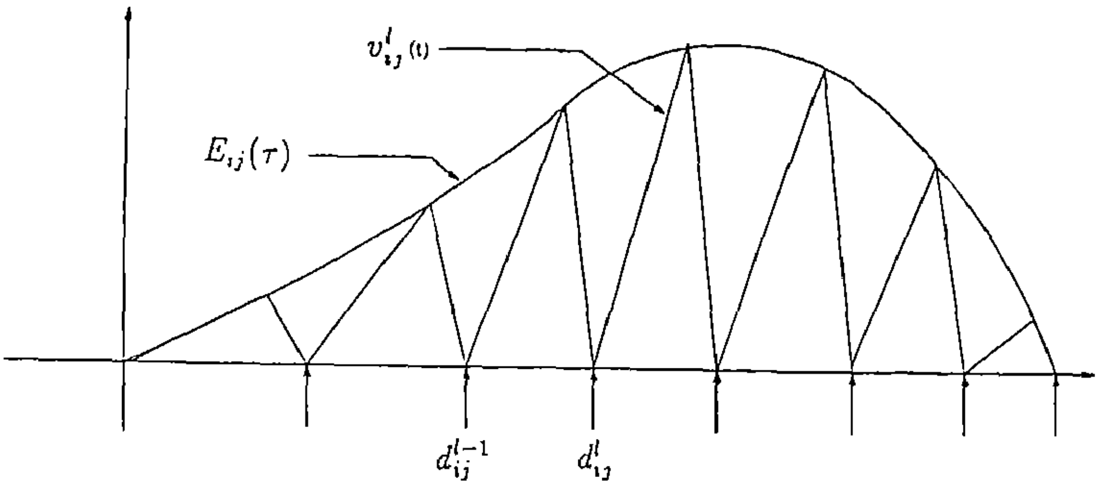


Figure 6.2: Non-uniform arrival pattern

done to make valid comparisons between the results of these cases.

6.3.2 Distribution of Buses

The two distribution of buses used are termed as *equal resource on each route* and *unequal resource on each route.*

Equal resource on each route

By equal resource, it means that at each station the number of buses on all the routes are the same. In this thesis, 10 buses on each of the routes at each of the stations are considered under this case.

Unequal resources on each route

As the name suggests, unequal resource on each route means unequal number of buses for different routes at a station. There are three routes at each of the stations and 8, 10 and 12 are assumed to ply on them.

6.3.3 Optimal Schedules

Schedules are developed for the following five cases on both the networks for equal and unequal resource on each route.

- Case I : Only Transfer time (TT); independent of arrival pattern of passengers.

- Case II : Only Initial waiting time (IWT); uniform arrival pattern.
- Case III : Total Waiting time (TWT); uniform arrival pattern.
- Case IV : Only Initial waiting time (IWT); non-uniform arrival pattern.
- Case V : Total Waiting time (TWT); non-uniform arrival pattern.

6.3.3.1 Optimal Schedules for Network - I

The first test network used is given in the Figure 2.2. In the network, there are two connecting routes; one between the Stations 1 and 2 and the other between Stations 2 and 3. The remaining four routes are isolated (i.e. they do not connect any two of the three stations in the network). The assumed travel times on the connecting routes are given in Section 6.2. The time period for which schedule is developed is constant at all the stations and is equal to 240 minutes, but there is a shift in schedule start times between two stations. This shift is equal to the travel time between the two stations plus the stopping time on the connecting route at the first station. For example, if at the first station the schedule start time and end times are *a* and *b* minutes respectively, the schedule start time and end times at the second station are $a+(c+d)$ and $b+(c+d)$ minutes respectively, where c is the travel time on the connecting route between the two stations and d is the stopping time of buses on the route at the first station. The stopping time is included in the shift to make the travel time for all the buses between two stations is same on the route, otherwise the last bus will have a travel time different from others as the last bus arrival time on a route is fixed at the end of the scheduling period.

The shifting of the schedule time is done due to the reason that if a bus on the common route arrives at the Station 1 at the earliest possible time, i.e. at time a minutes, then it can arrive at the Station 2 only after a period of $c+d$ minutes. Thus the bus can arrive only at $a+c+d$ minutes. In this way, all the buses on the common route between Stations 1 and 2 arrive at Station 2 $c+d$ minutes after their arrival at Station 1. In this problem, all the buses at the first station are within the scheduling period 0 to 240 minutes. All the buses at Station 2 are within 30 plus stopping time of Route 1 (common route) at Station 1 and 270 plus stopping time of Route 1 at Station 1. Similarly all the buses at Station 3 are within 70 plus stopping time of Route 5 at Station 2 and 310 plus stopping time of Route 5 at Station 2.

Equal resource on each route

The schedules for equal resource on each route with uniform as well as non-uniform arrival pattern of passengers are given below. When the number of buses are same on all the routes at a station and the arrival pattern of passengers is uniform, then it is possible to predict an optimal solution. This is explained during the discussion of the results next. The following results are presented for the five cases defined earlier.

Case - I

In this case, schedules based on only transfer time consideration are presented; i.e. schedules are independent of the nature of the arrival pattern of passengers. For minimum transfer time of passengers, the corresponding buses should arrive at the same time on all the routes at a station since number of buses are same on all the routes. And the stopping times on the routes should be minimum for the optimal transfer time for passengers. There can be infinite number of optimal solutions for this case, since any schedule in which arrival times are same on different routes at a station is an optimal solution.

The optimal schedule based on only transfer time consideration is given in the Figures 6.3(a),6.3(b) and 6.3(c). Each figure presents the schedule at one of the three stations. On the top of each figure, station number, the generation in which this schedule is obtained, number of buses on each of the routes at the station and the objective function value are given. In the figures, the horizontal axis represents the time in minutes, and the route numbers at the station are given on the vertical axis. In all the figures, routes at bottom are isolated with respect to other stations, the middle routes are the routes which connect to other stations and the top routes are those which come other stations if any. The arrow pointing to the horizontal line represents the arrival of a bus and the arrow emanating from it represents the departure.

As seen from the figures, the corresponding buses on different routes at each station arrive at almost the same time. The total transfer time for

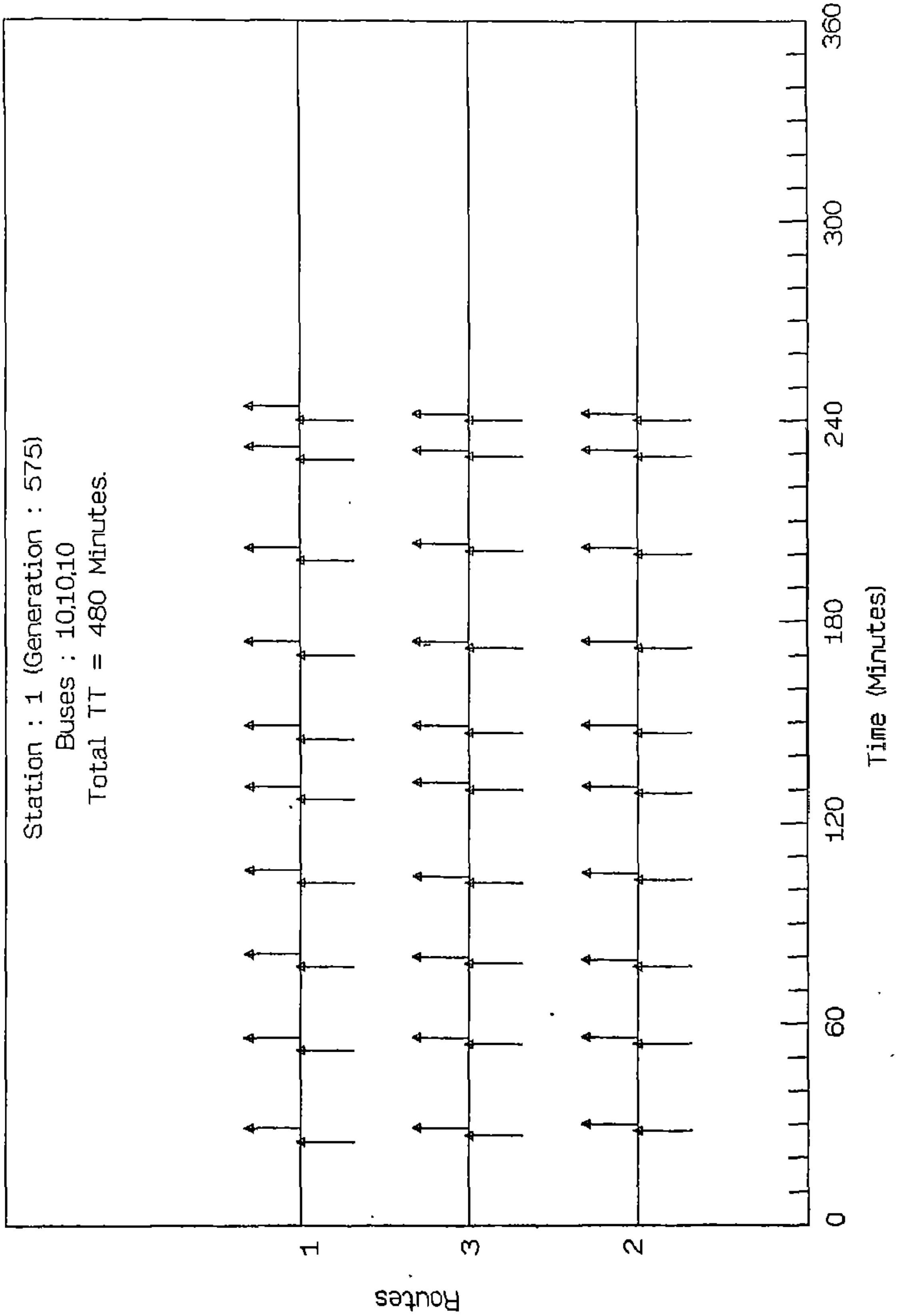


Figure 6.3(a): Optimal schedule for TT consideration

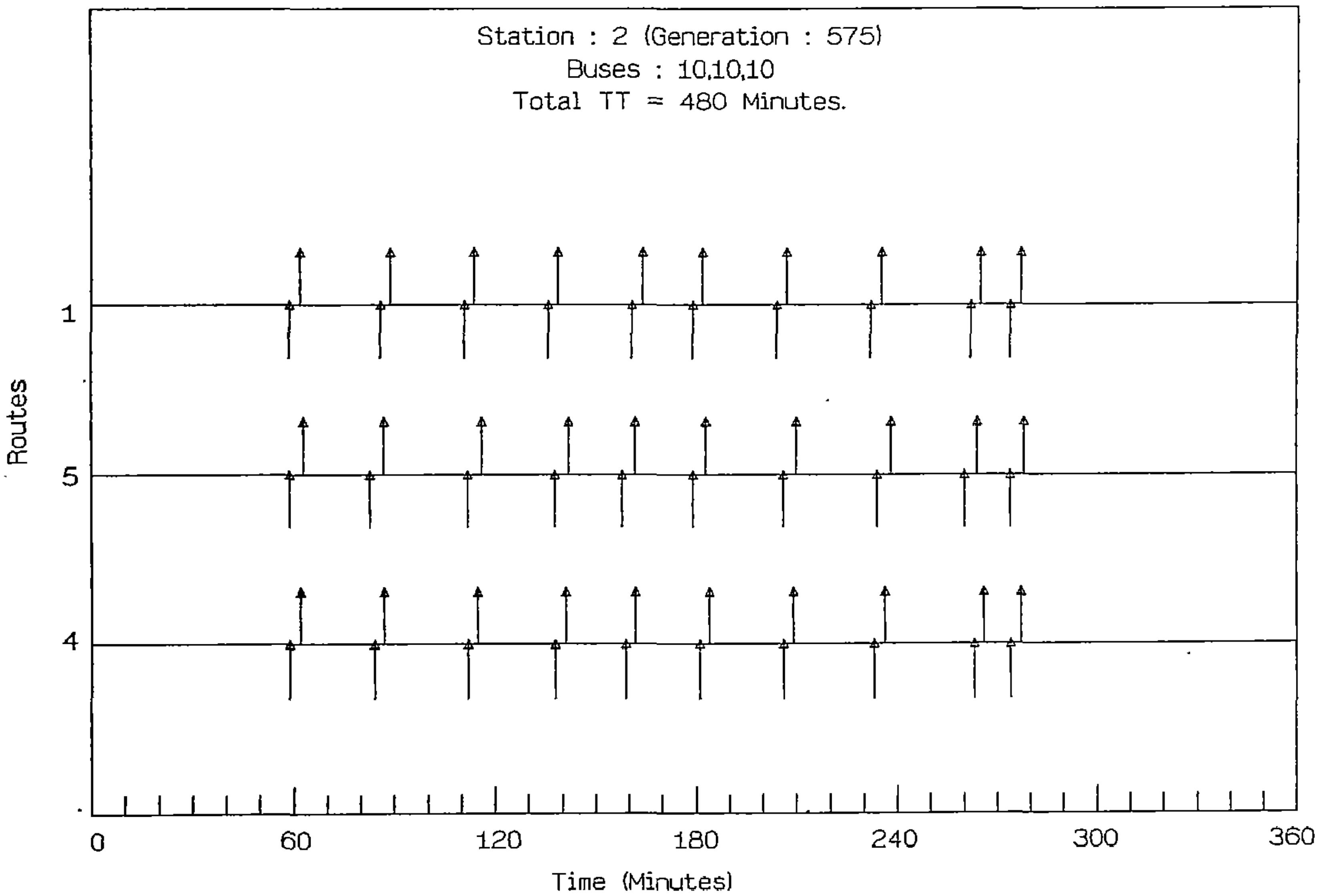


Figure 6.3(b): Optimal schedule for TT consideration

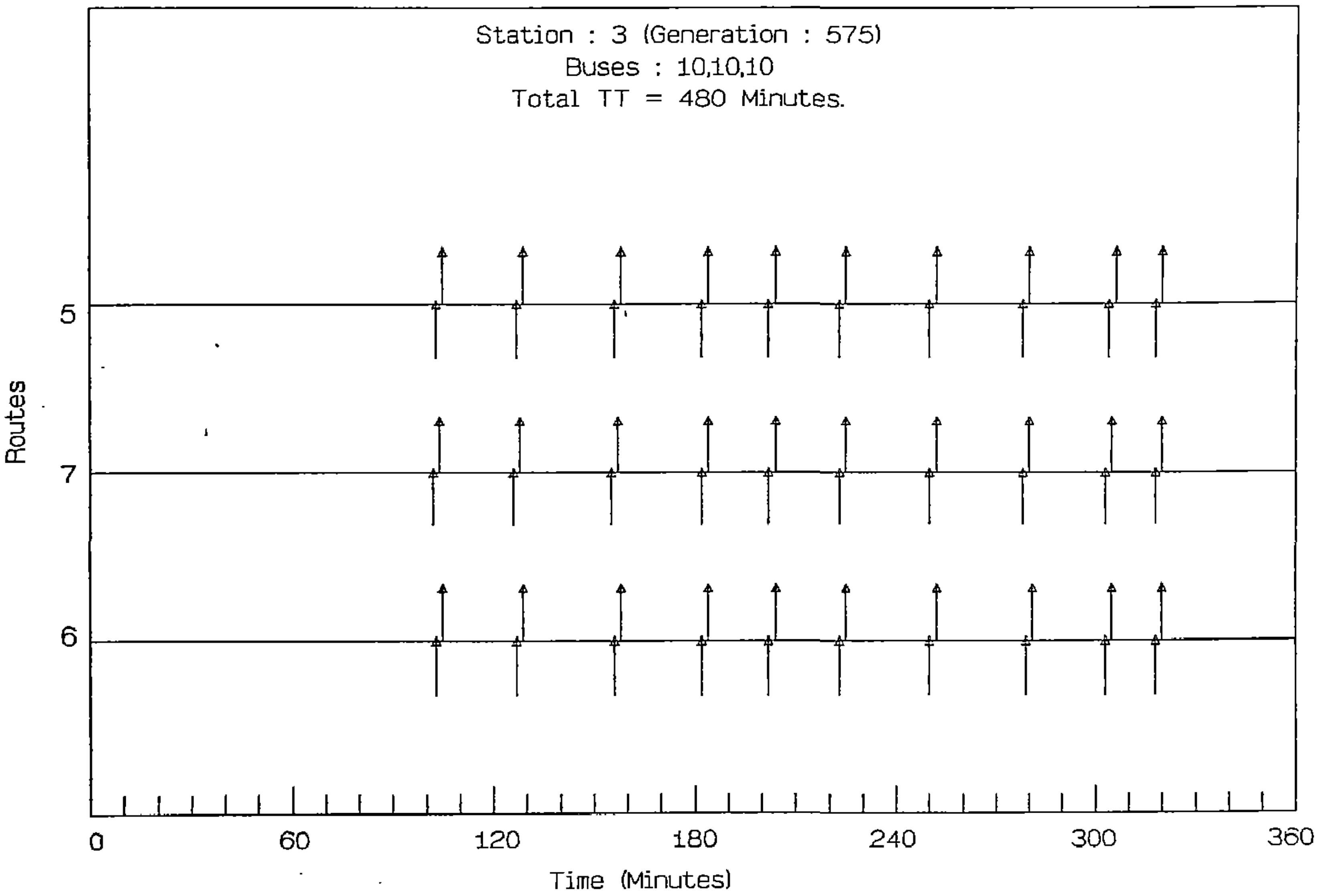


Figure 6.3(c): Optimal schedule for TT consideration

the schedule is 480 minutes which is 120 minutes more than the minimum possible transfer time of 360 minutes. The minimum possible transfer time is obtained when the corresponding buses on different routes at a station arrive at the same time and stopping time on all the routes is equal to minimum stopping time of 2 minutes. In the schedule shown in the figures, the stopping time on some routes is more than 2 minutes, that is why the transfer time is more than the minimum possible. However, the purpose of this case is to show that using GAs the basic optimality condition-that corresponding buses on each route should arrive at the same time-is achieved.

Case - II

When only initial waiting time is considered with uniform arrival pattern of passengers, at the optimal the buses should arrive at uniform headway on each route (As seen from the objective function, the initial waiting time term is in the form of Σx_i^2, where x_i represents the headway between consecutive buses on a route. Minimizing $\Sigma_{i=1}^{n} x_i^2$ subject to $\Sigma_{i=1}^{n} x_i = k$ gives the solution $x_i = \frac{k}{n}$, where k is the scheduling period and n is the total number of buses on the route. So the headways should be equal for minimum initial waiting time).

The schedules obtained for initial waiting time consideration are given in the Figures 6.4(a), 6.4(b) and 6.4(c). As seen from the figures, the buses for each route at each station arrive with uniform headways. The total initial waiting time obtained is 15570 minutes which is only 4 percent more than the absolute minimum. It may be mentioned again, that GAs achieved the basic optimality criterion (uniform headways) in this case also.

Case - III

Total initial waiting time is the sum of the transfer time and initial waiting time of passengers. In this case also we can expect uniform headways since the solution which is optimal for IWT case is also optimal for TT case. The schedule for this case is shown in Figures 6.5(a), 6.5(b) and 6.5(c). As expected, the headways are uniform. This result may be compared with results of Case I to see the effect of including IWT in the objective function. The total waiting time obtained is 16078 minutes which is approximately 5

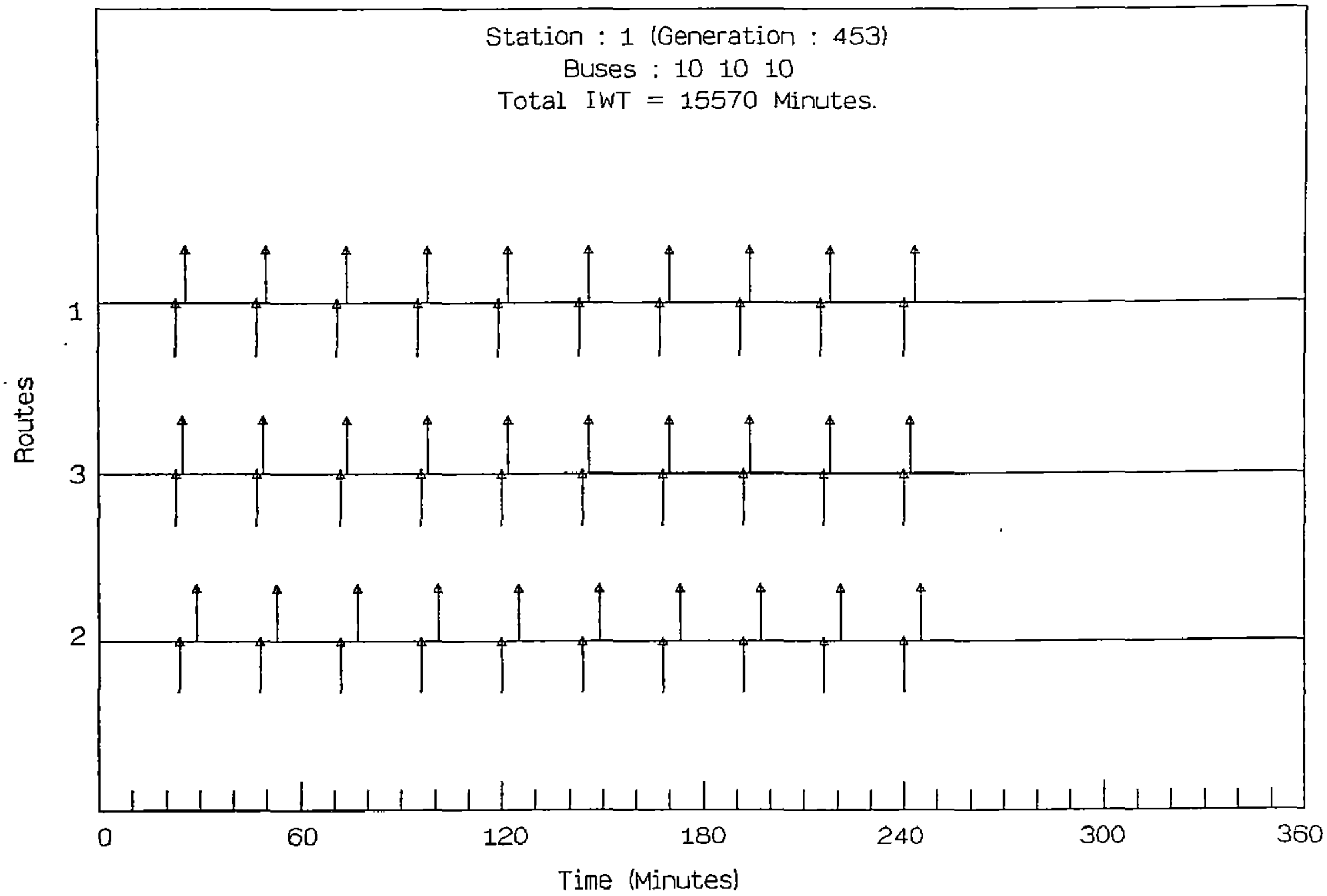


Figure 6.4(a): Optimal schedule for IWT consideration (Uniform arrival pattern)

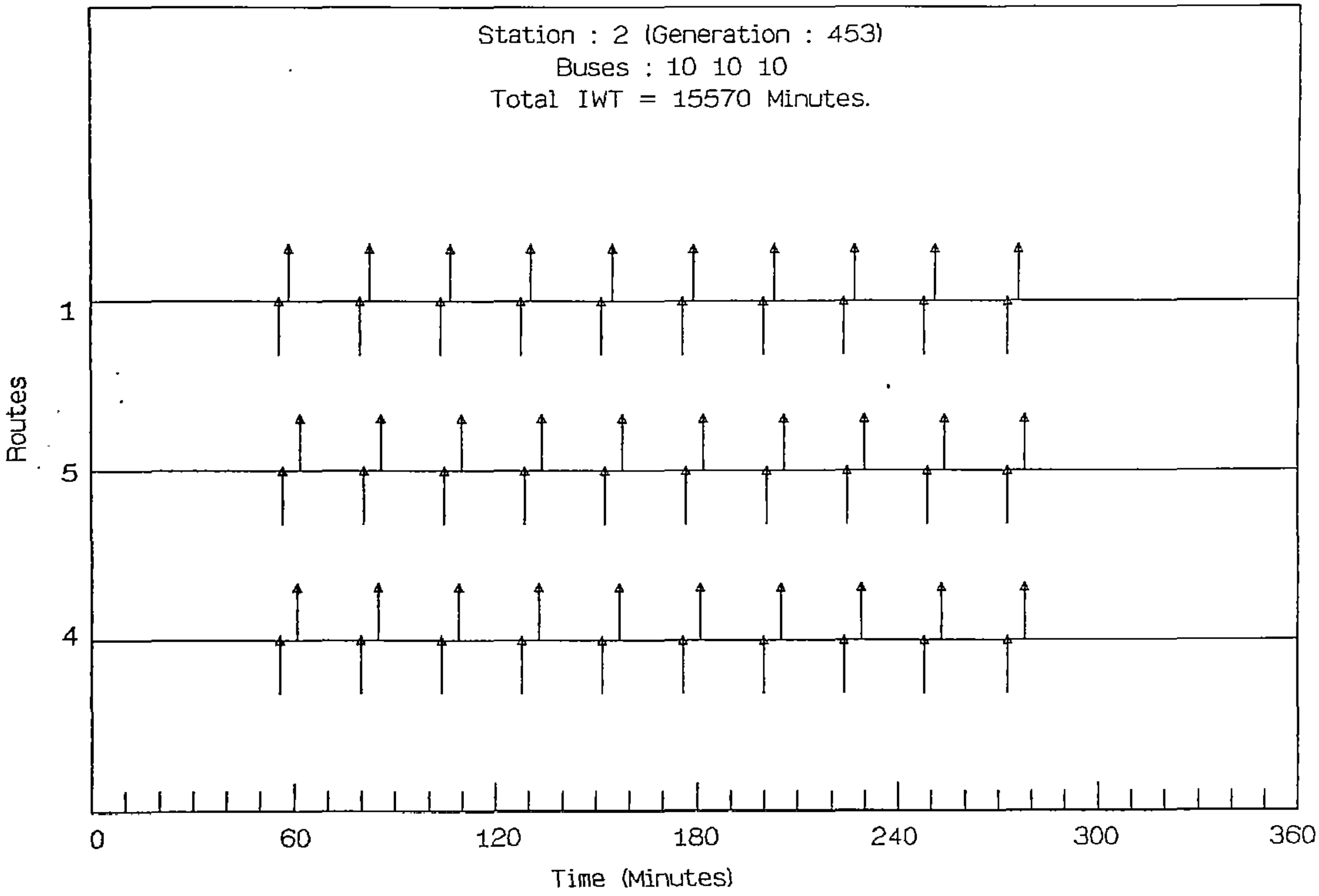


Figure 6.4(b): Optimal schedule for IWT consideration (Uniform arrival pattern)

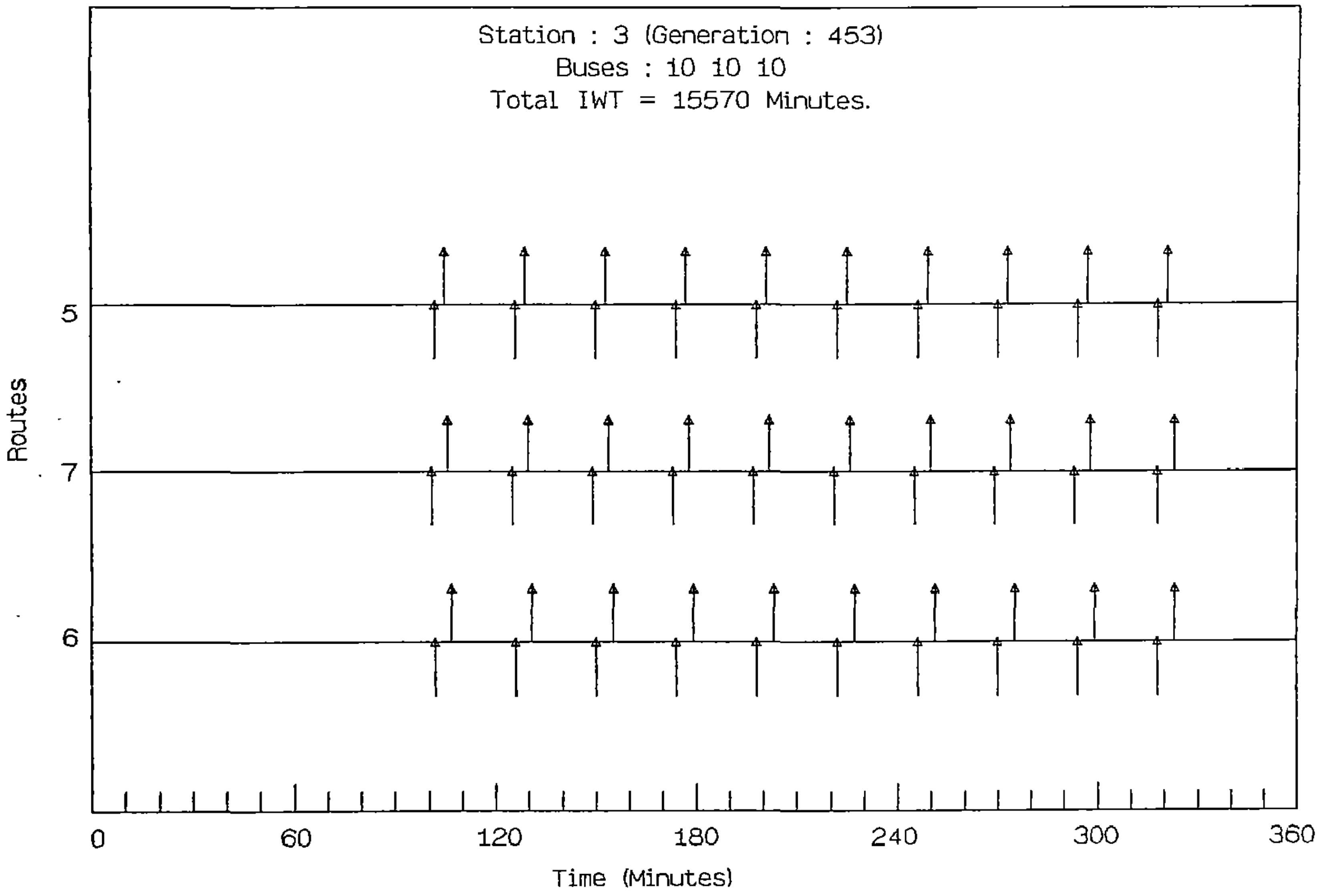


Figure 6.4(c): Optimal schedule for IWT consideration (Uniform arrival pattern)

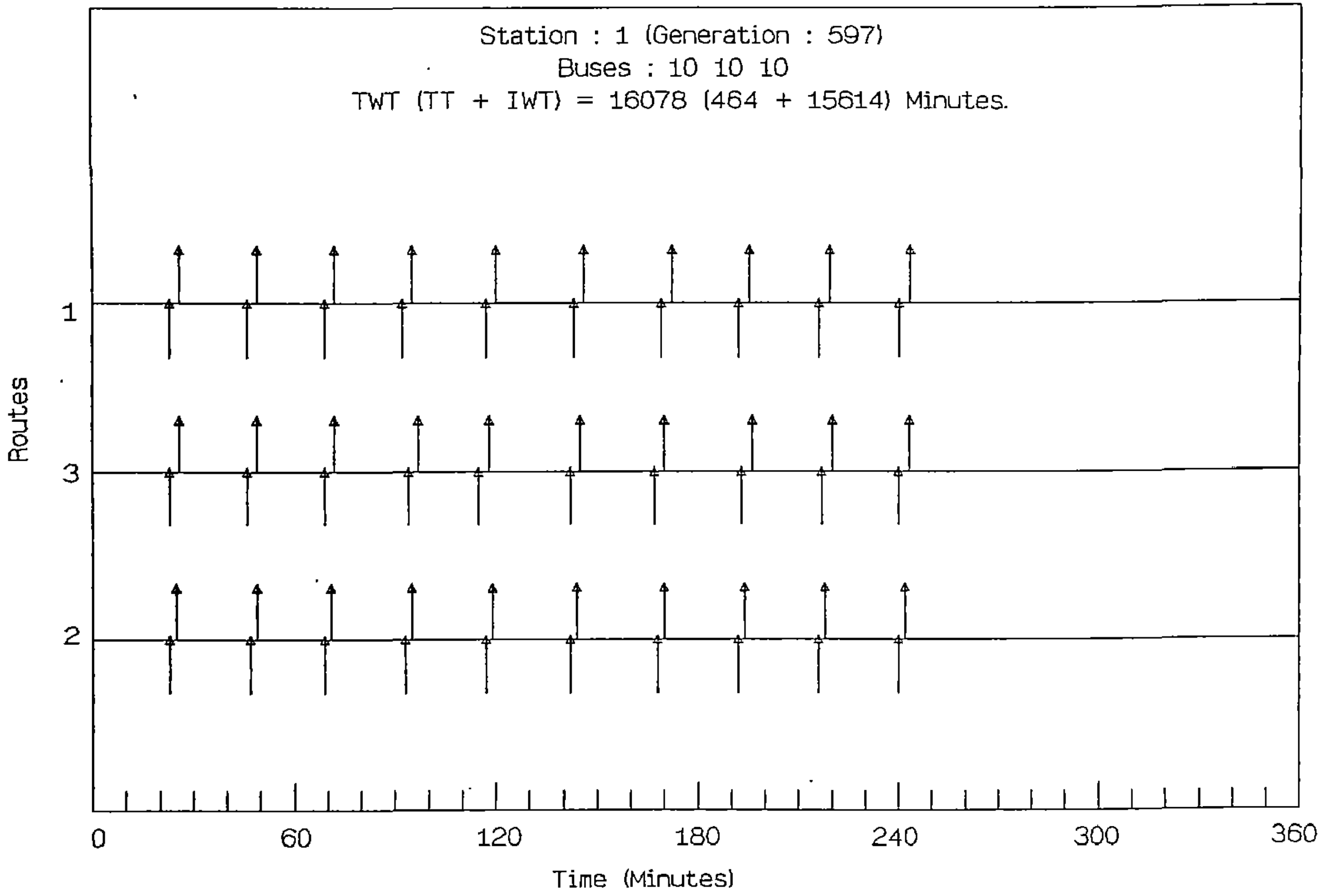


Figure 6.5(a): Optimal schedule for TWT consideration (Uniform arrival pattern)

CENTRAL LIBRARY
I. I. T., KANPUR
Acc. No. A. 121391

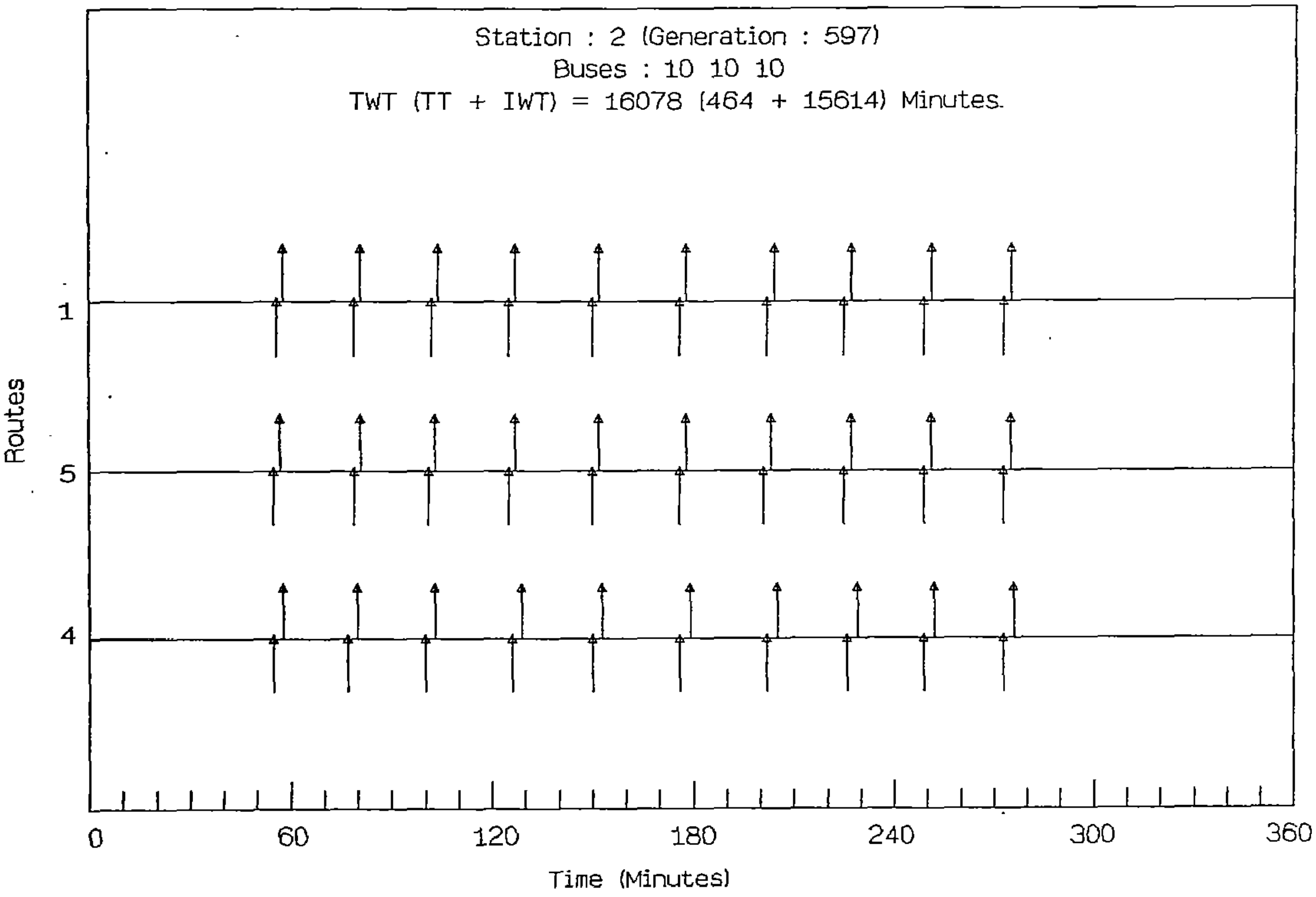


Figure 6.5(*b*): Optimal schedule for TWT consideration (Uniform arrival pattern)

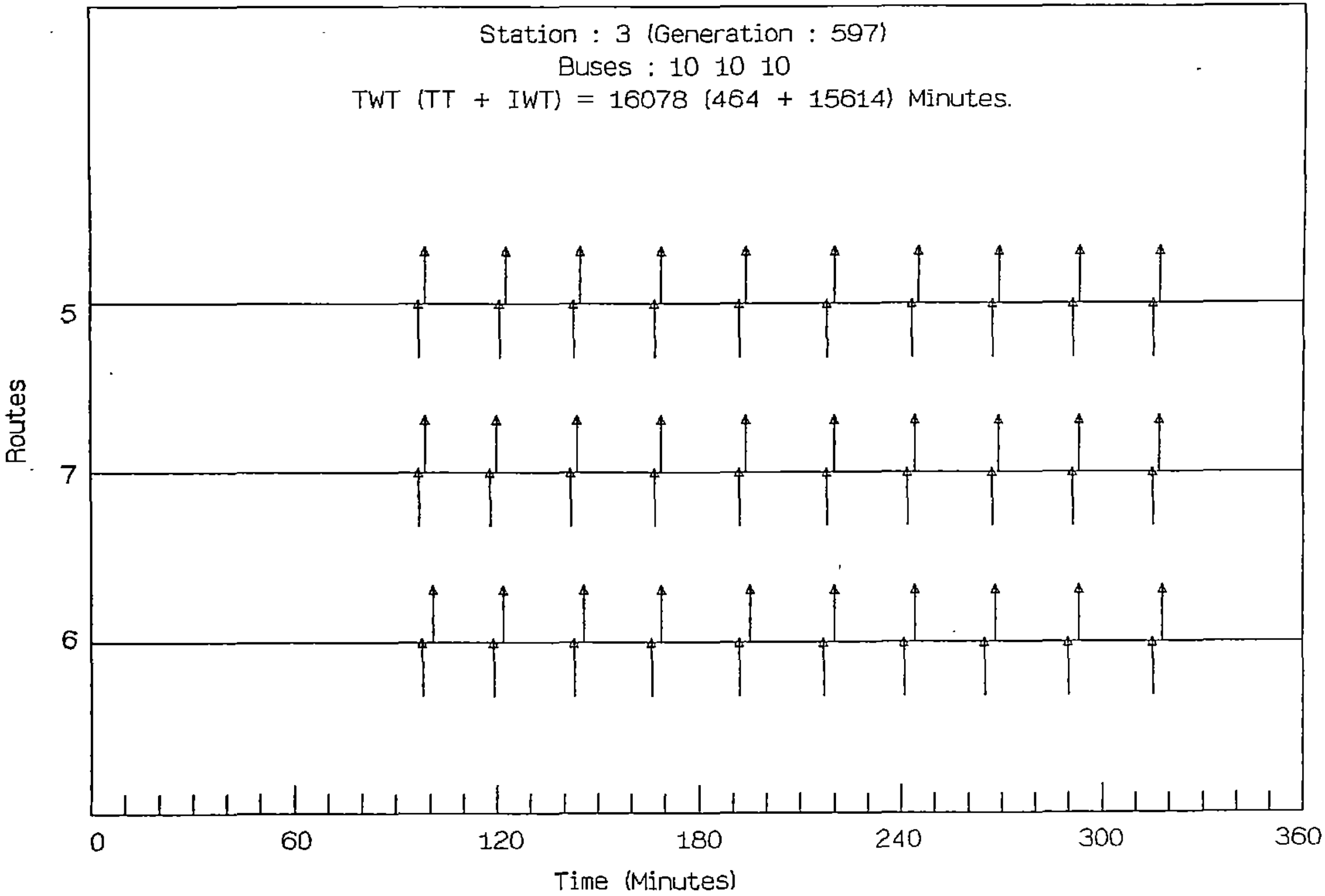


Figure 6.5(c): Optimal schedule for TWT consideration (Uniform arrival pattern)

percent more than the absolute minimum for this case. The headways are uniform at all the stations and the arrival times coincide for corresponding buses at a station.

Case - IV

The optimal schedule based on only initial waiting time consideration for non-uniform arrival pattern of passengers is given in the Figures 6.6(a), 6.6(b) and 6.6(c). The total initial waiting time is 15056 minutes. As more number of passengers come in the latter half of the scheduling period, more buses arrive in that half at all the stations. So the headways are less in the second half when compared with first half of the scheduling period. The last headway is larger than the rest in the second half of the scheduling period because less number of passengers arrive during this period.

Case - V

In this case (i.e. with TWT consideration) as in the case of initial waiting time consideration, one expects more buses in the second half of the scheduling period at all the stations. The optimal schedule for this case is shown in the Figures 6.7(a), 6.7(b) and 6.7(c). The total waiting time is 15740 minutes. As seen from the figures, more buses consequently with less headway come in the second half of the scheduling period since more number of passengers arrive in this half.

As seen from the schedules presented thus far, the results are promising because (1) they approach the absolute minimum values for uniform arrival pattern and (2) show predictive changes in the schedules for non-uniform arrival patterns. This shows that the genetic algorithms work efficiently in solving the scheduling problem. Next, the optimal schedules for unequal resources on each route are presented.

Unequal resources on each route

The optimal schedules obtained for unequal resources on each route at all stations with uniform as well as non-uniform arrival pattern of passengers are presented in this section.

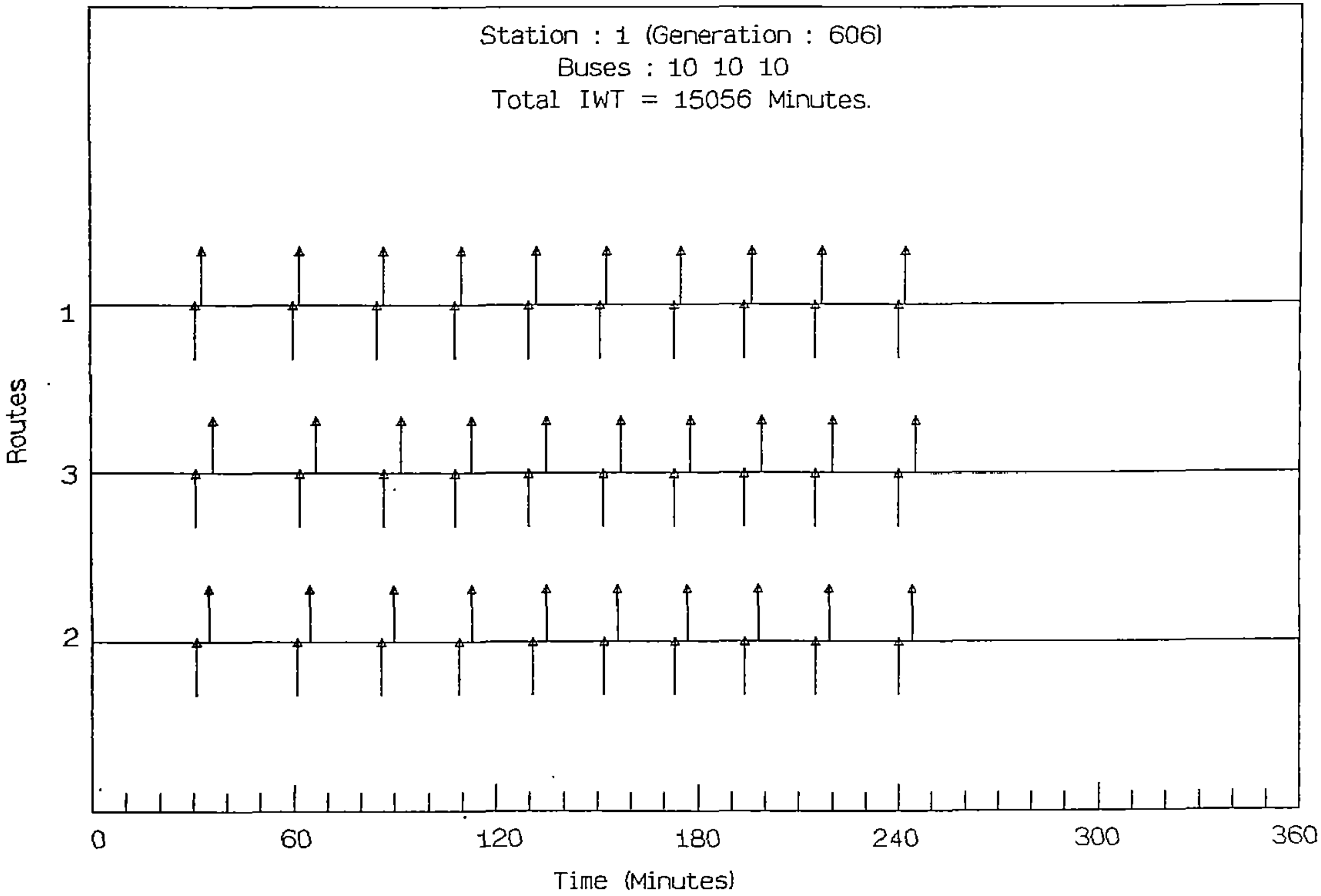


Figure 6.6(a): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

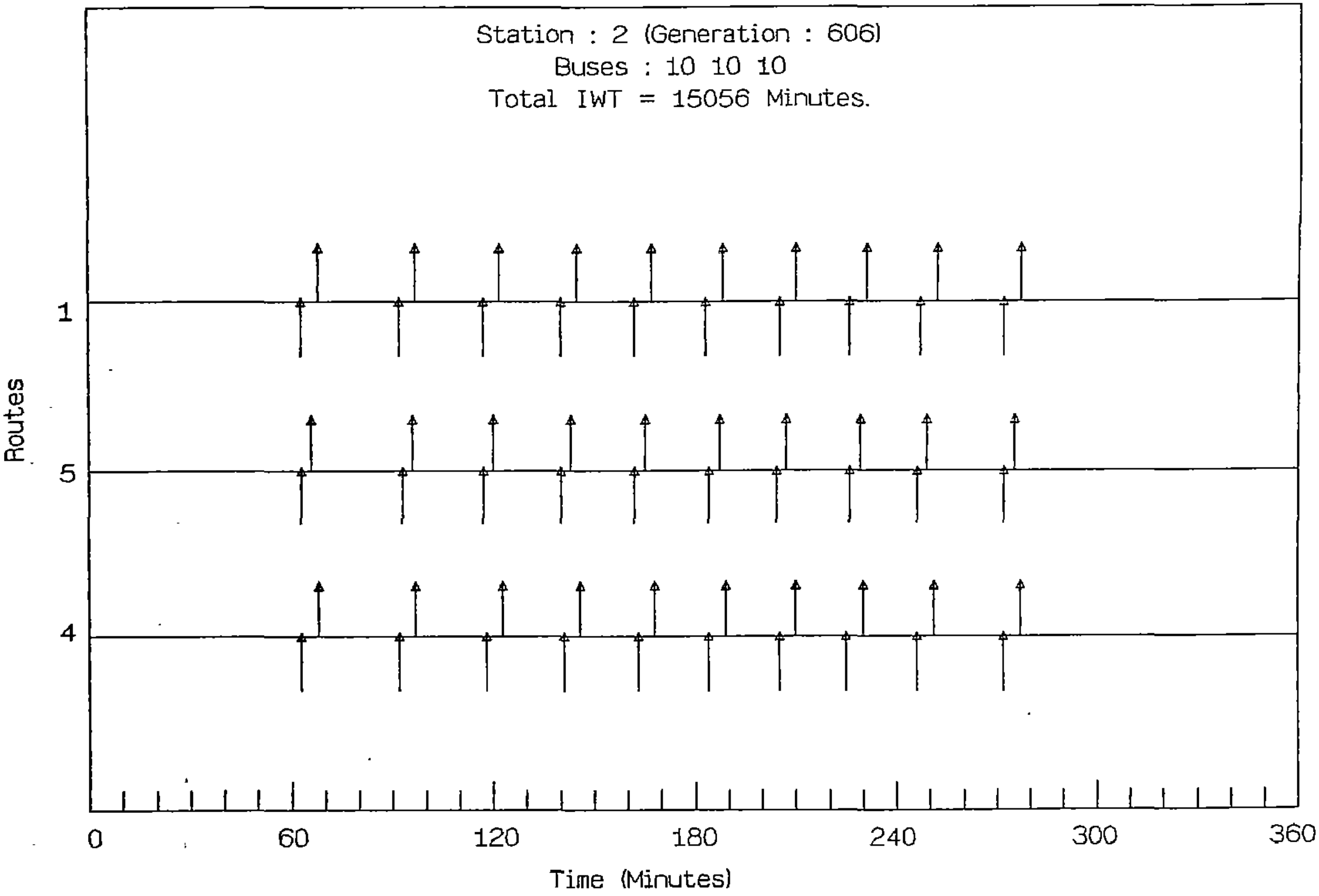


Figure 6.6(*b*): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

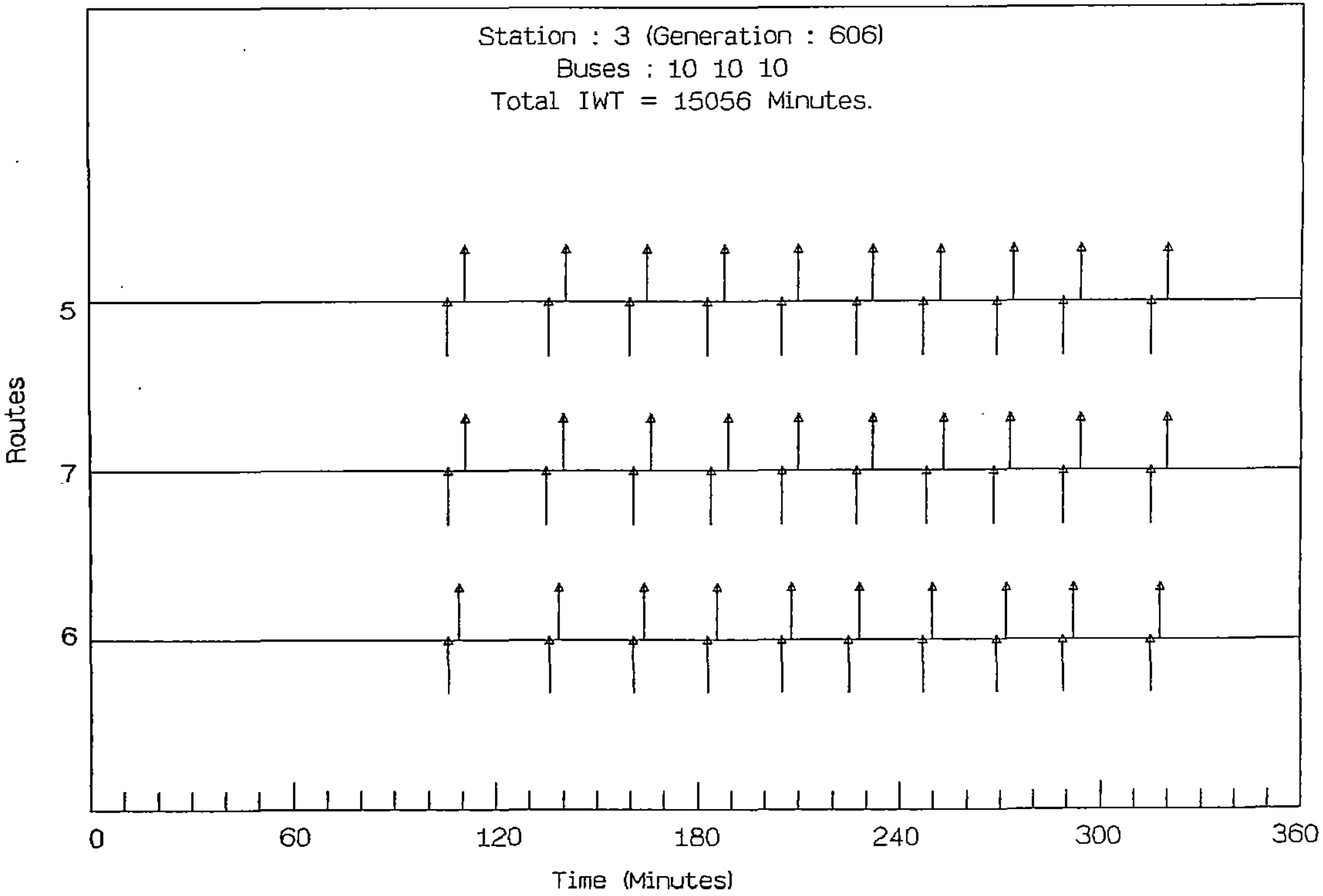


Figure 6.6(c): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

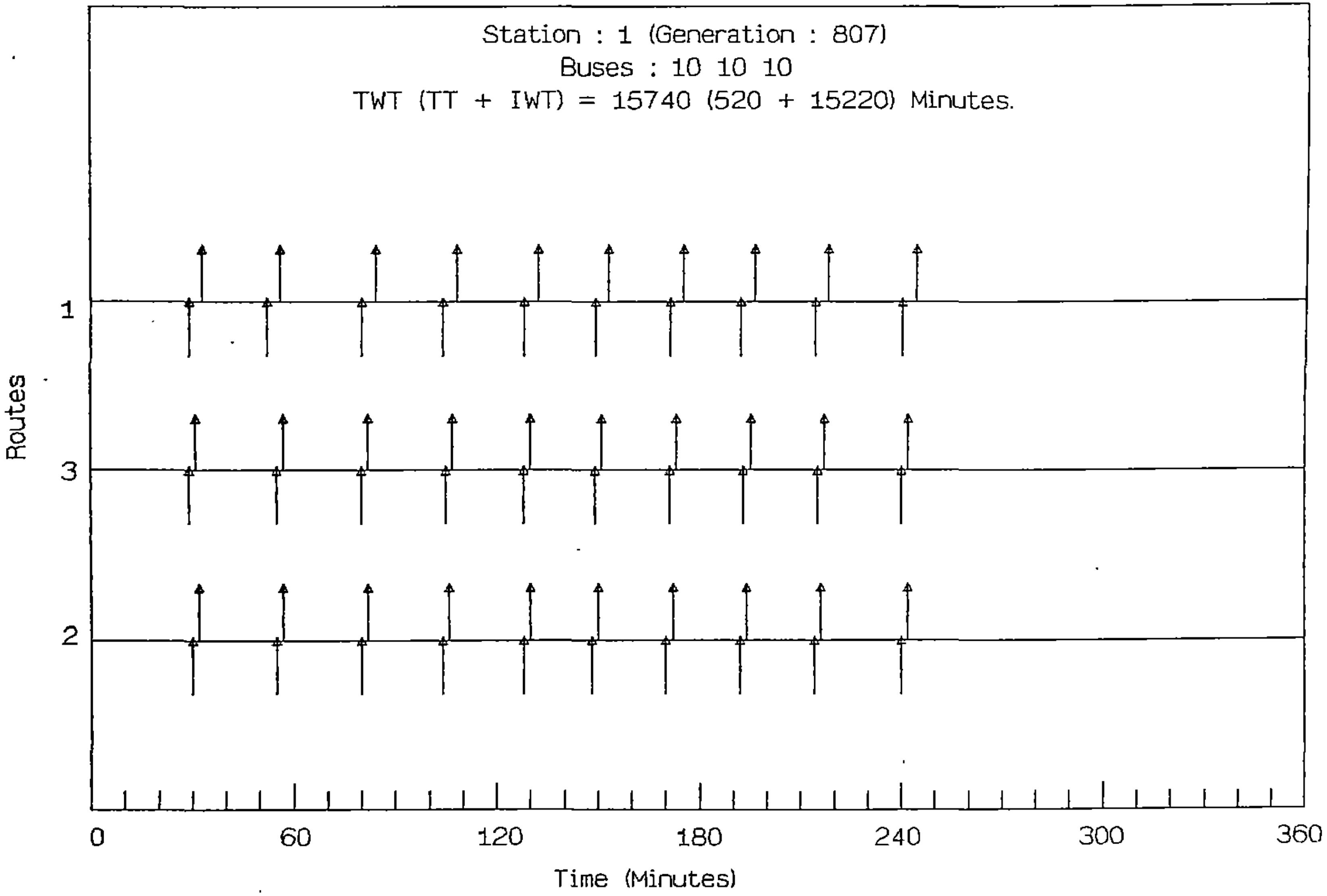


Figure 6.7(a): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

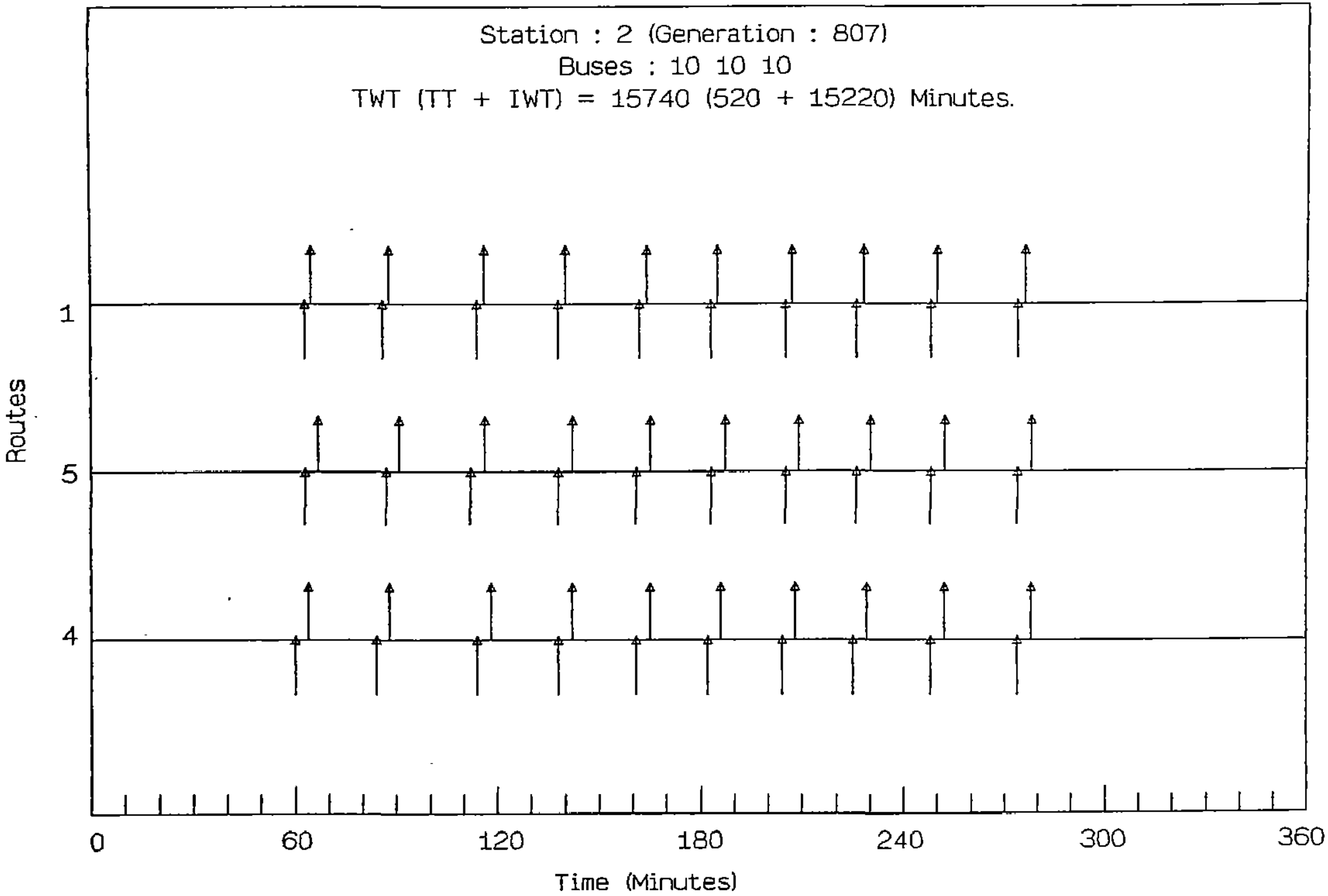


Figure 6.7(*b*): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

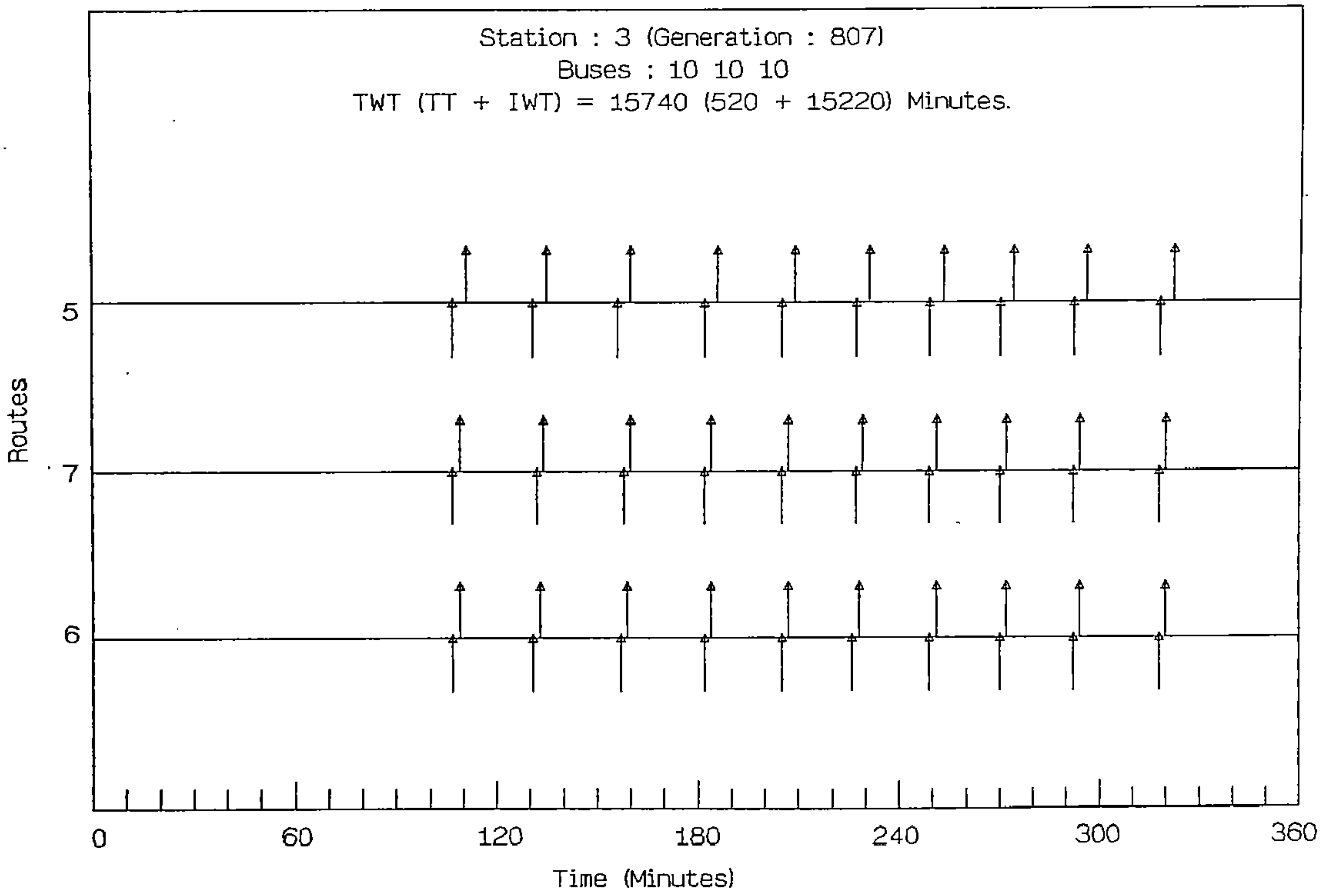


Figure 6.7(c): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

Case - I

The optimal schedule obtained is shown in the Figures 6.8(a), 6.8(b) and 6.8(c). The total transfer time obtained is 833 minutes. As given earlier, for only TT consideration under equal resource the corresponding buses on all the routes at a station should arrive at the same time. However, in the present case the number of buses are not same on all the routes at a station, hence the corresponding buses should not come at the same time. As seen from the schedule, the buses align themselves in a manner such that no two buses on different routes are too far. The buses try to adjust in such a way so that transfer time is minimum.

Case - II

The optimal schedule based on only initial waiting time consideration for uniform arrival pattern of passengers is given in the Figures 6.9(a), 6.9(b) and 6.9(c). The total initial waiting time obtained is 15576 minutes. For only IWT consideration with uniform arrival pattern of passengers, the optimal schedule would be with uniform headways on all the routes. The schedule obtained is supports this fact. The buses on all the routes at each station arrive at a uniform headway.

Case - III

When both the transfer time and initial waiting time are considered, we cannot expect any predicatable schedule as in the case of equal resource on each route since in this case, the number of buses are not same on all the routes at a station. The optimal schedule obtained in this case is shown in the Figures 6.10(a), 6.10(b) and 6.10(c). This schedule is obtained under the assumption that the number of passengers transferring from one route to another for each transfer is one. The total TT, according to this schedule is 1340 minutes and the total waiting time is 17153 minutes.

As opposed to the corresponding case for equal resource on each route, the optimal solution from the IWT stand point is suboptimal from the TT consideration. Thus, one can now investigate the changes in the schedule with variations in w_{ijk}^{l}. Note that this investigation is not possible in case III for equal resource on each route because the optimal condition for IWT

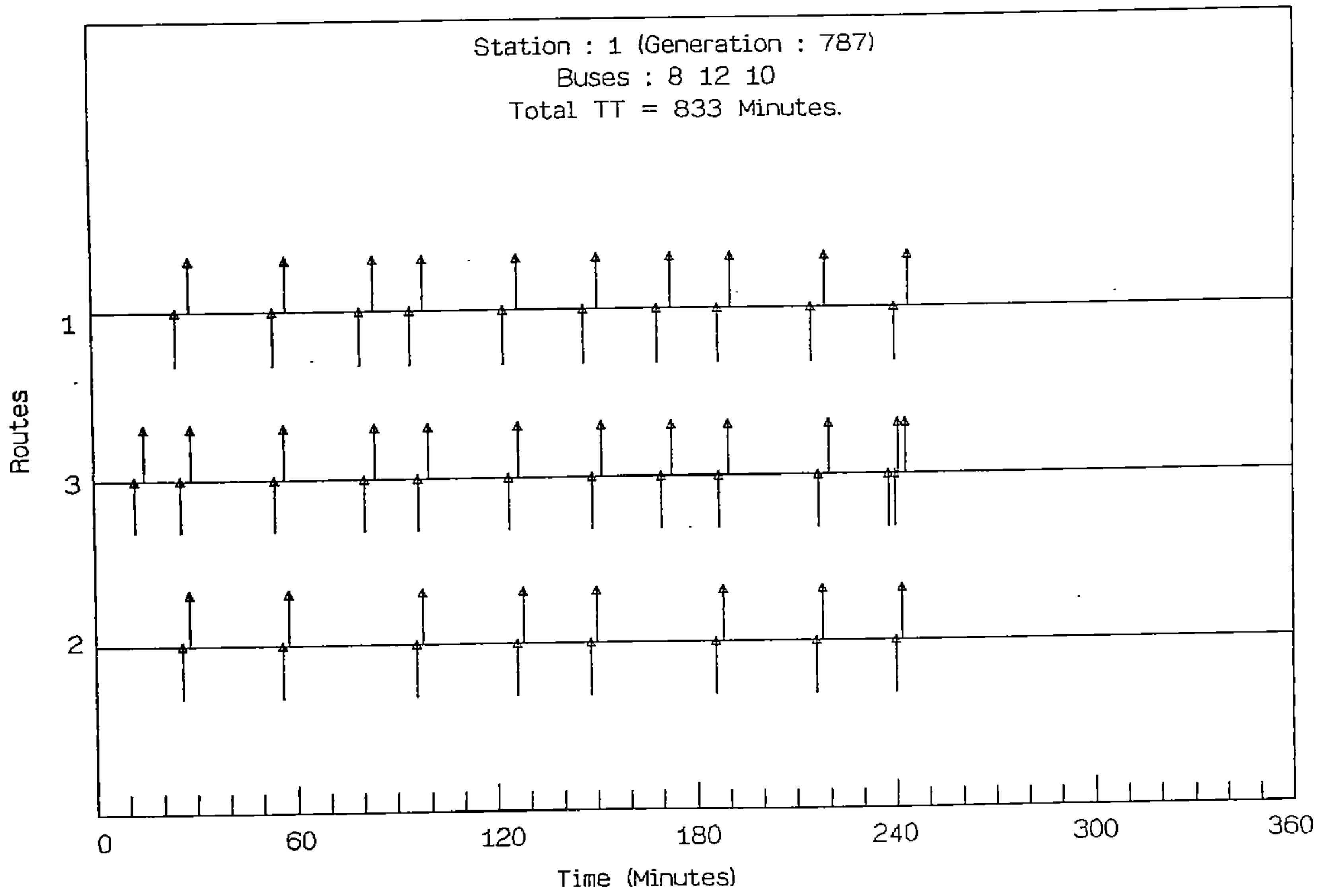


Figure 6.8(a): Optimal schedule for TT consideration

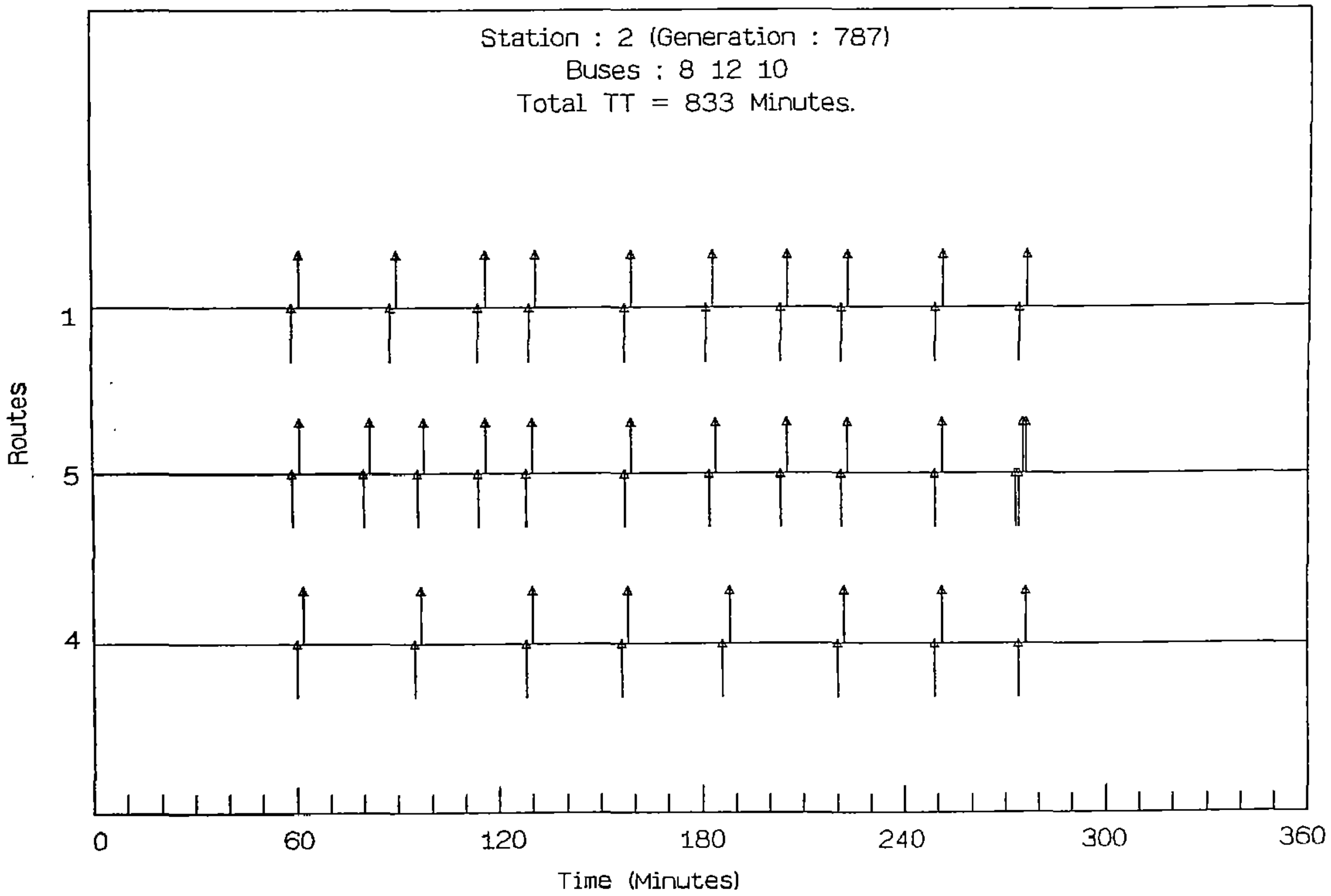


Figure 6.8(*b*): Optimal schedule for TT consideration

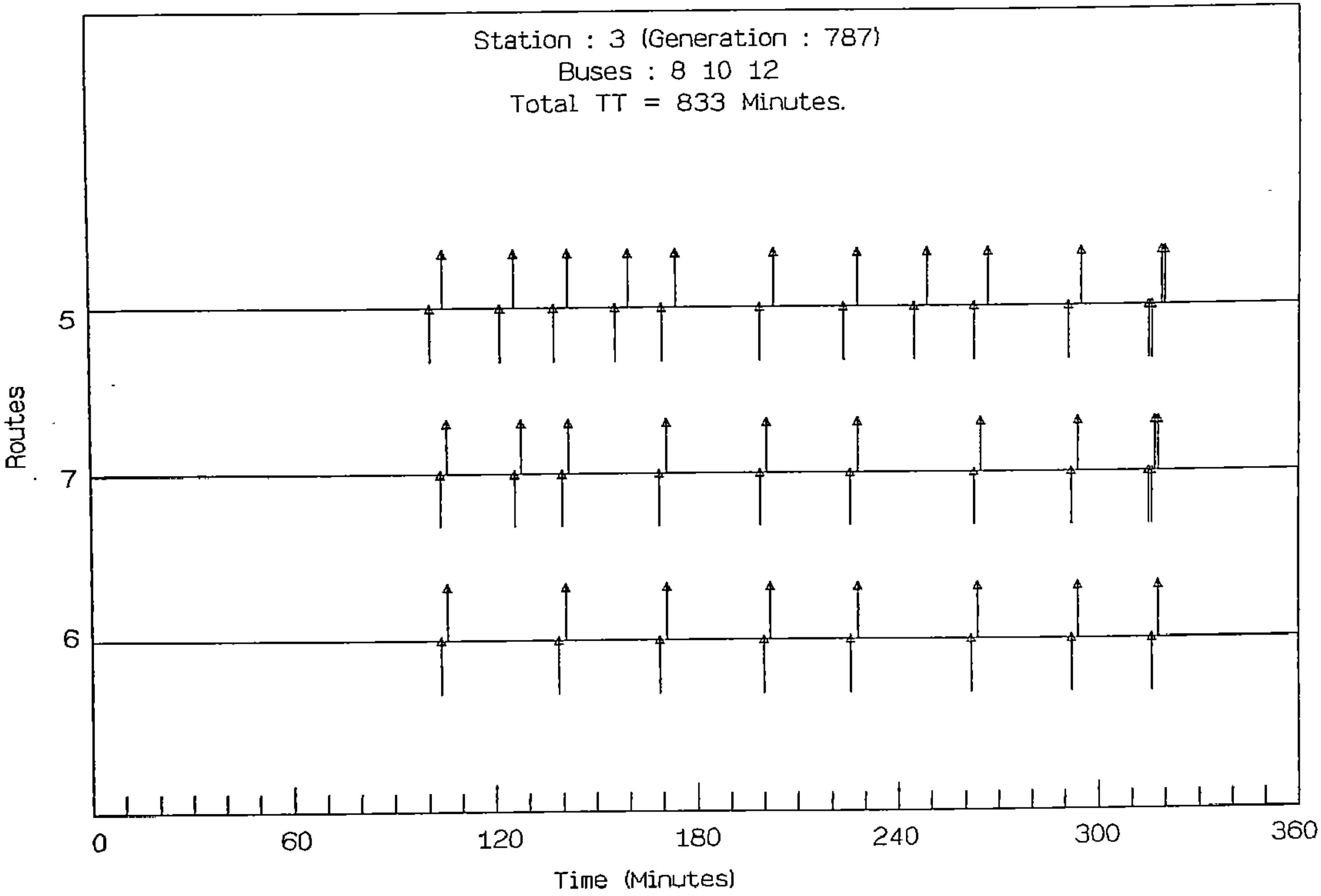


Figure 6.8(c): Optimal schedule for TT consideration

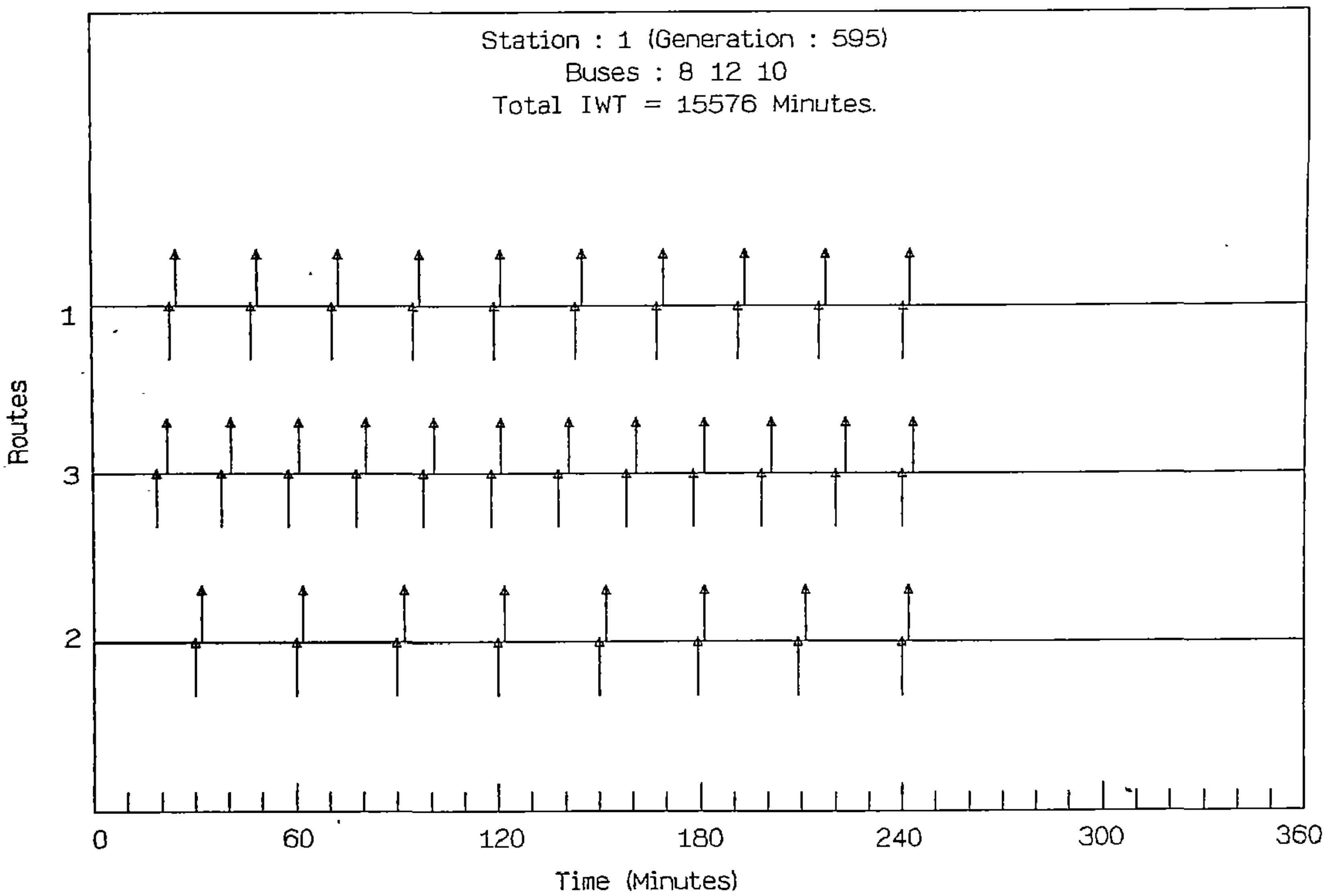


Figure 6.9(a): Optimal schedule for IWT consideration (Uniform arrival pattern)

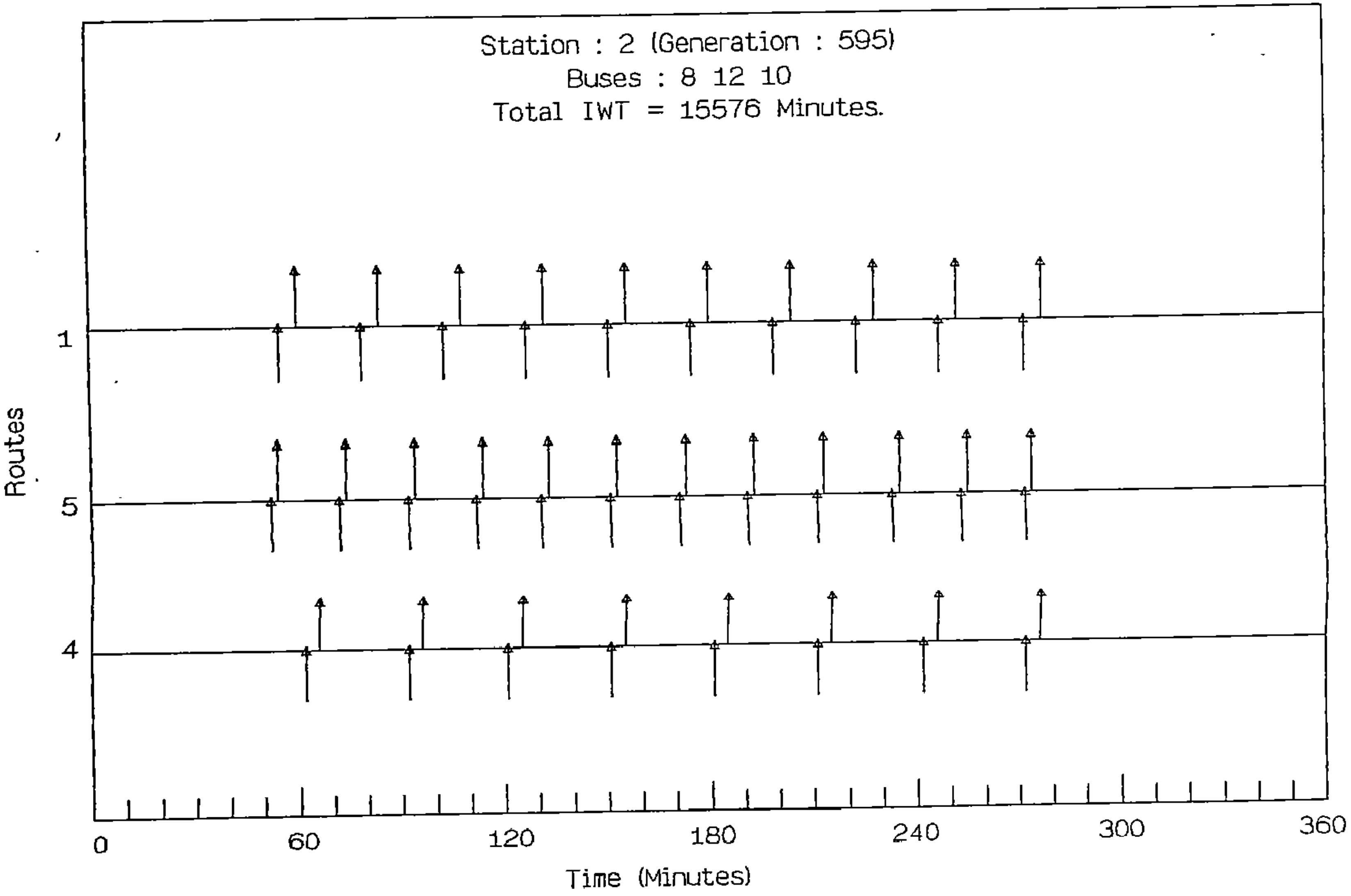


Figure 6.9(b): Optimal schedule for IWT consideration (Uniform arrival pattern)

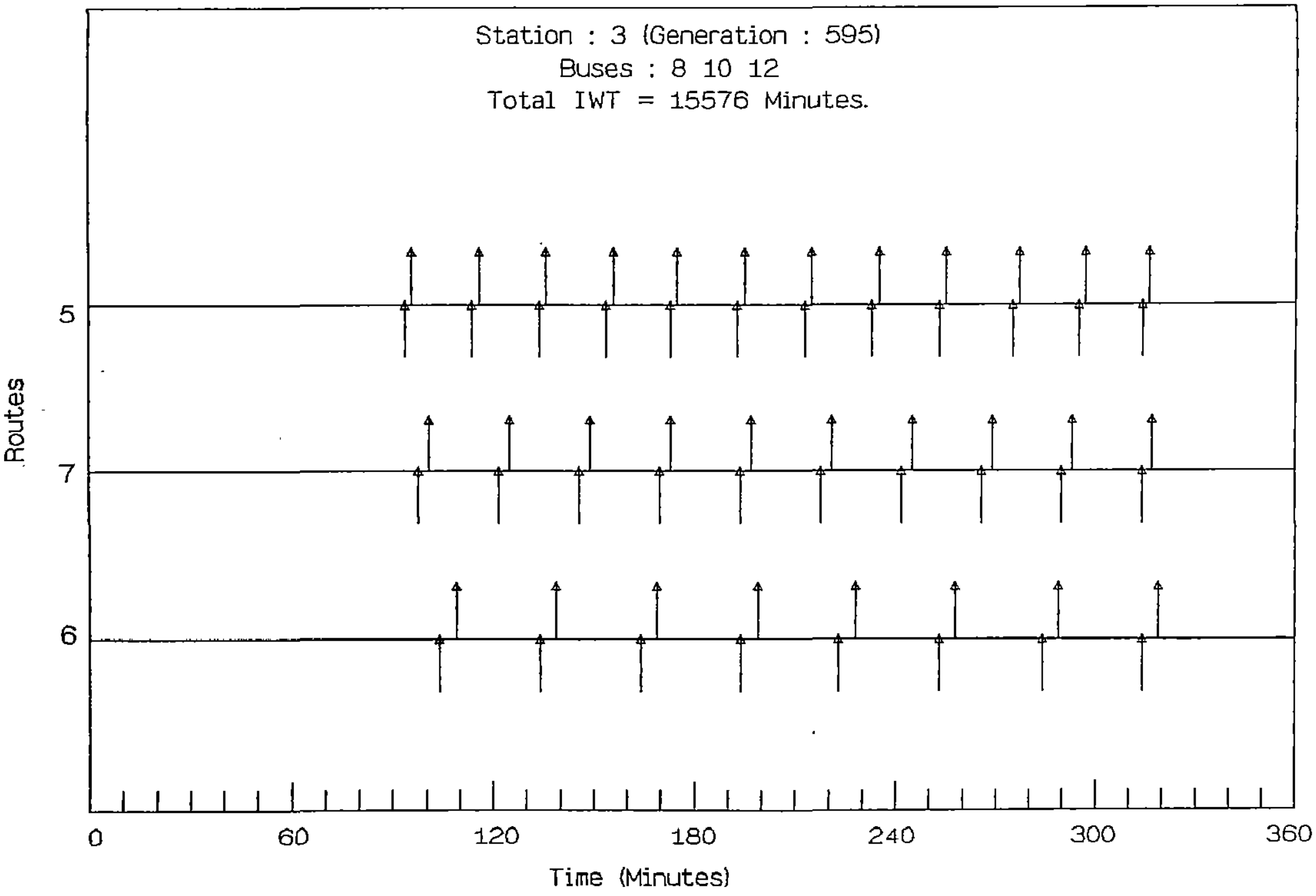


Figure 6.9(c): Optimal schedule for IWT consideration (Uniform arrival pattern)

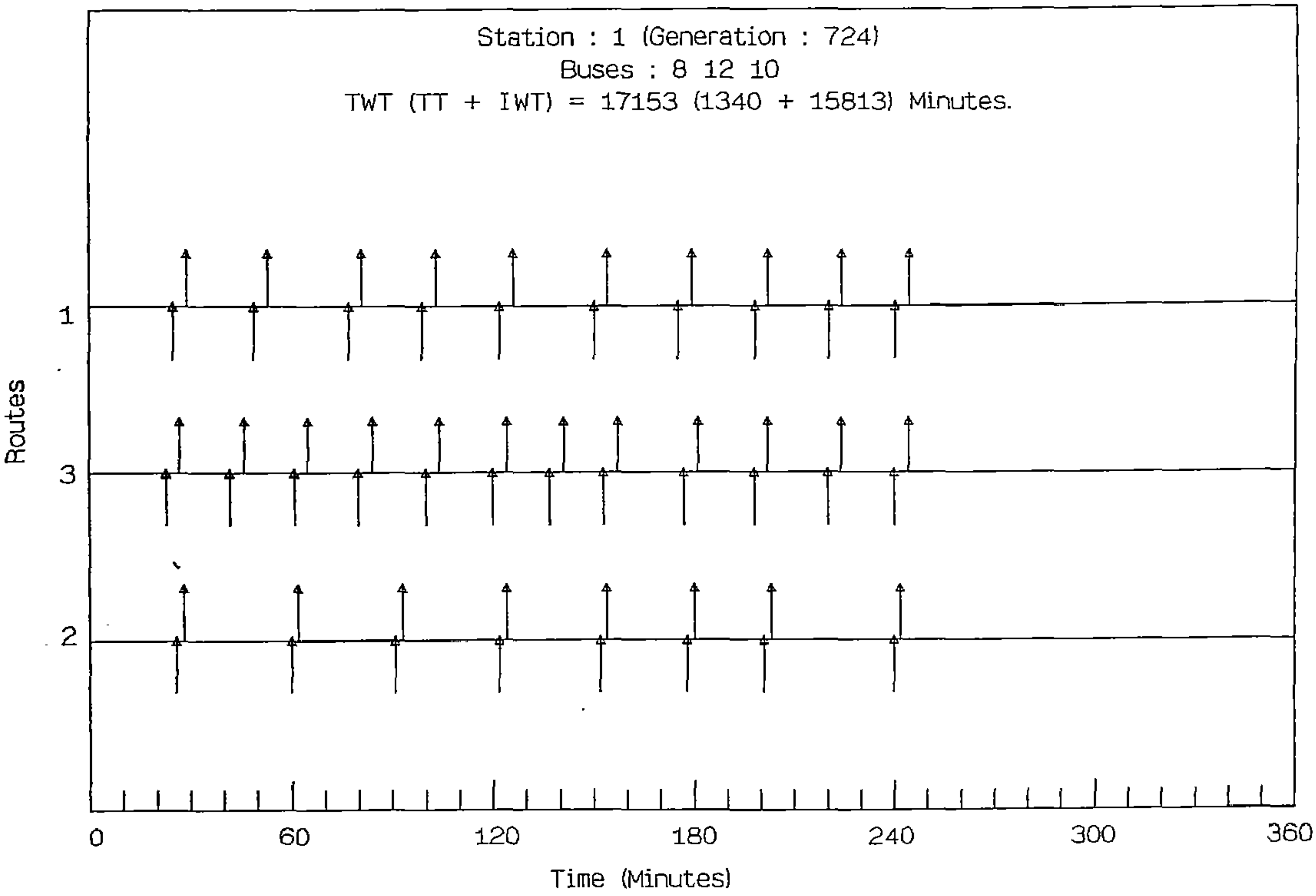


Figure 6.10(a): Optimal schedule for TWT consideration (Uniform arrival pattern)

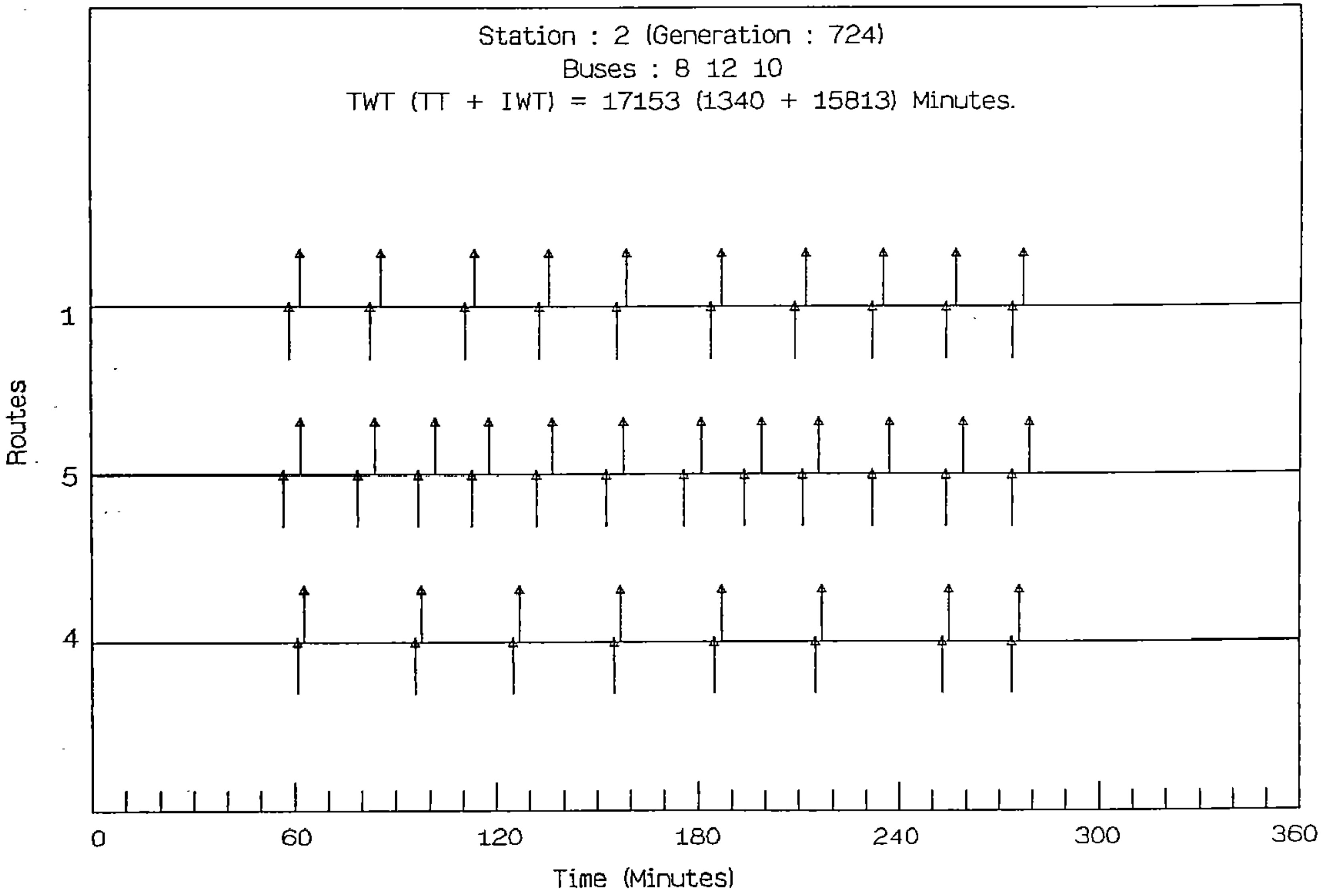


Figure 6.10(b): Optimal schedule for TWT consideration (Uniform arrival pattern)

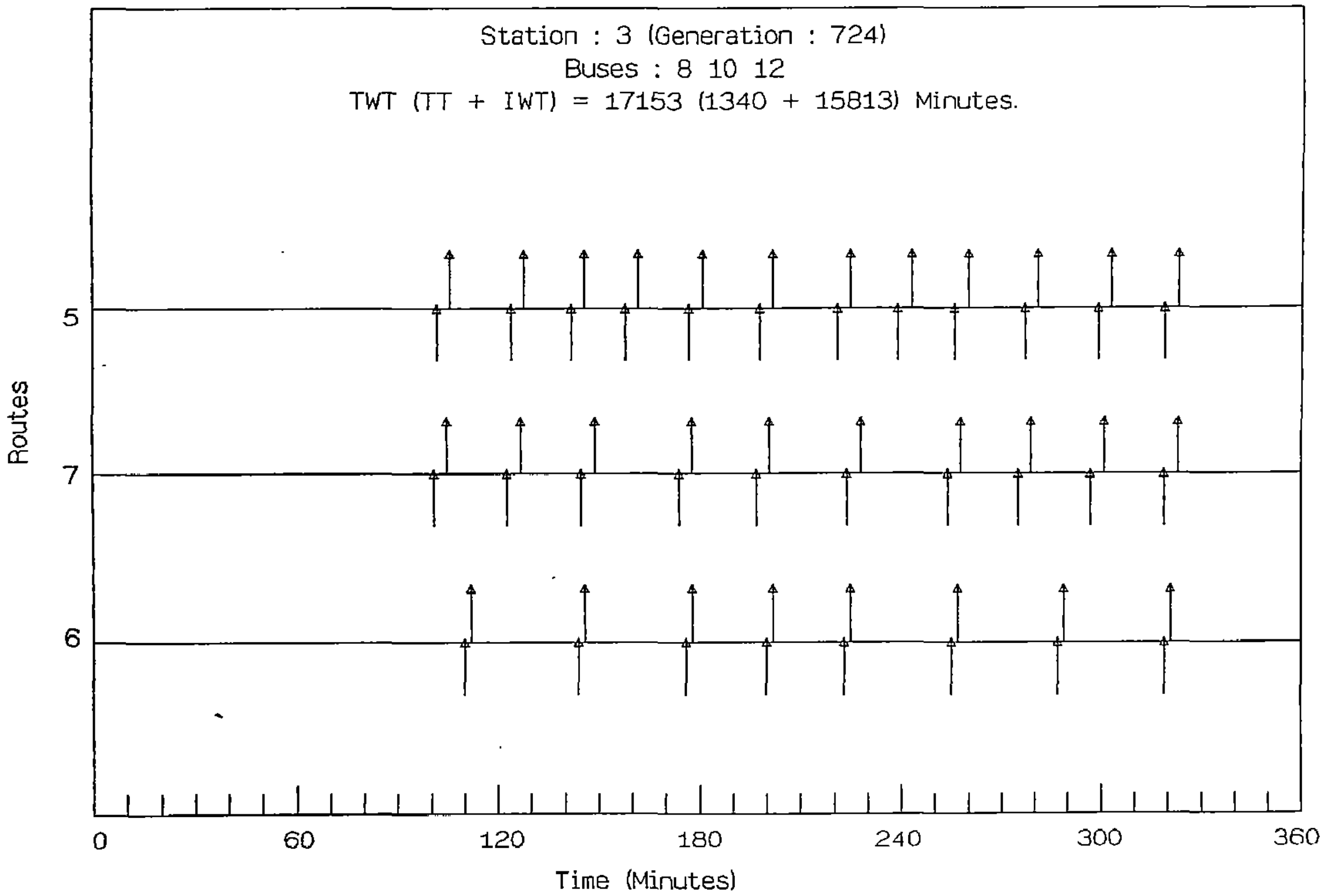


Figure 6.10(c): Optimal schedule for TWT consideration (Uniform arrival pattern)

consideration was also optimal for TT consideration.

One would expect that as the number of transferring passengers increase the schedule will modify itself to suit the transferring passengers more and more. That is the average transfer time per transferring passenger should reduce with increase in w^l_{ijk}. This fact can be seen from the Figure 6.11. The points in the figure are obtained from GA runs for different w^l_{ijk} values. The curve is a best fit line for these points.

Case - IV

The optimal schedule obtained for only IWT consideration with non-uniform arrival pattern of passengers is shown in the Figures 6.12(a), 6.12(b) and 6.12(c). In this case, as the peak of the β function defining the arrival pattern of passengers is on the second half of the scheduling period at all the stations, more number of passengers will come in the second half of the scheduling period. So, more buses should come in the second half for waiting

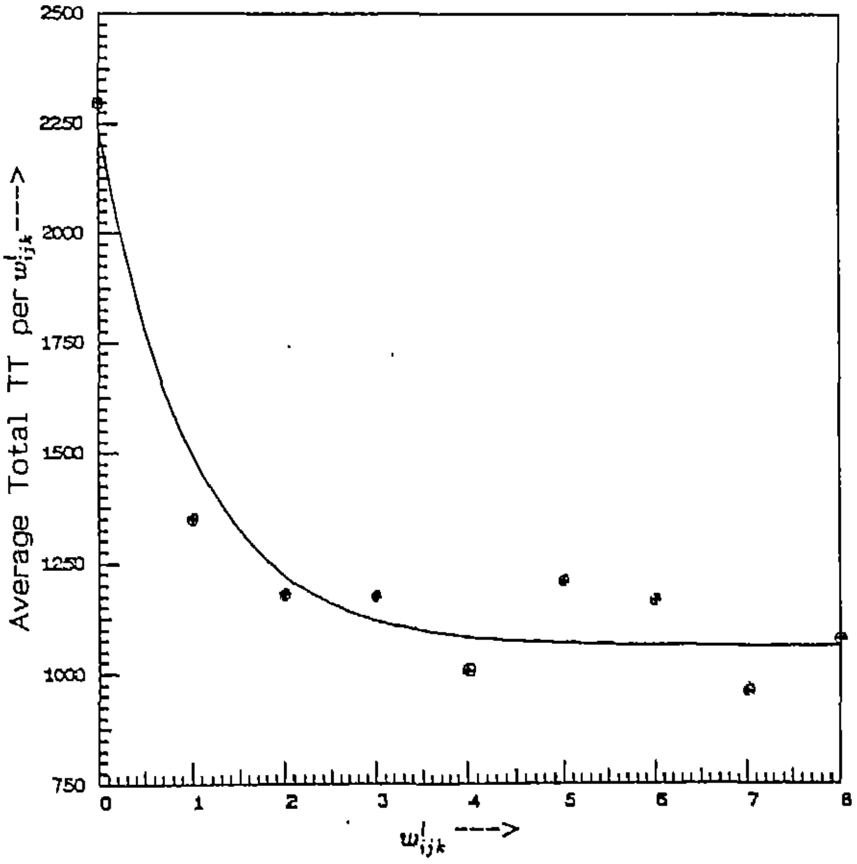


Figure 6.11: Variance of Average total TT per w^l_{ijk} with w^l_{ijk}.

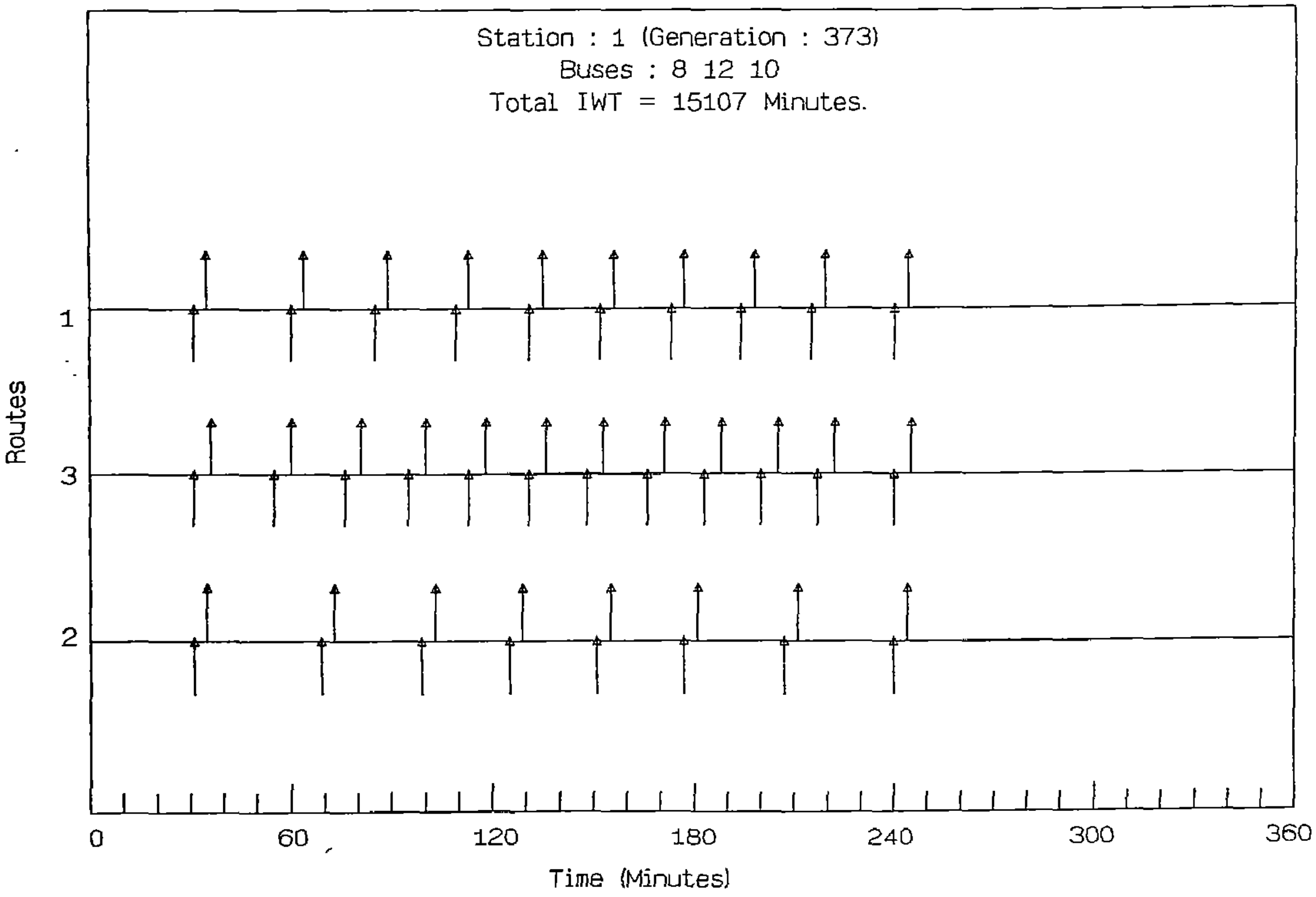


Figure 6.12(a): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

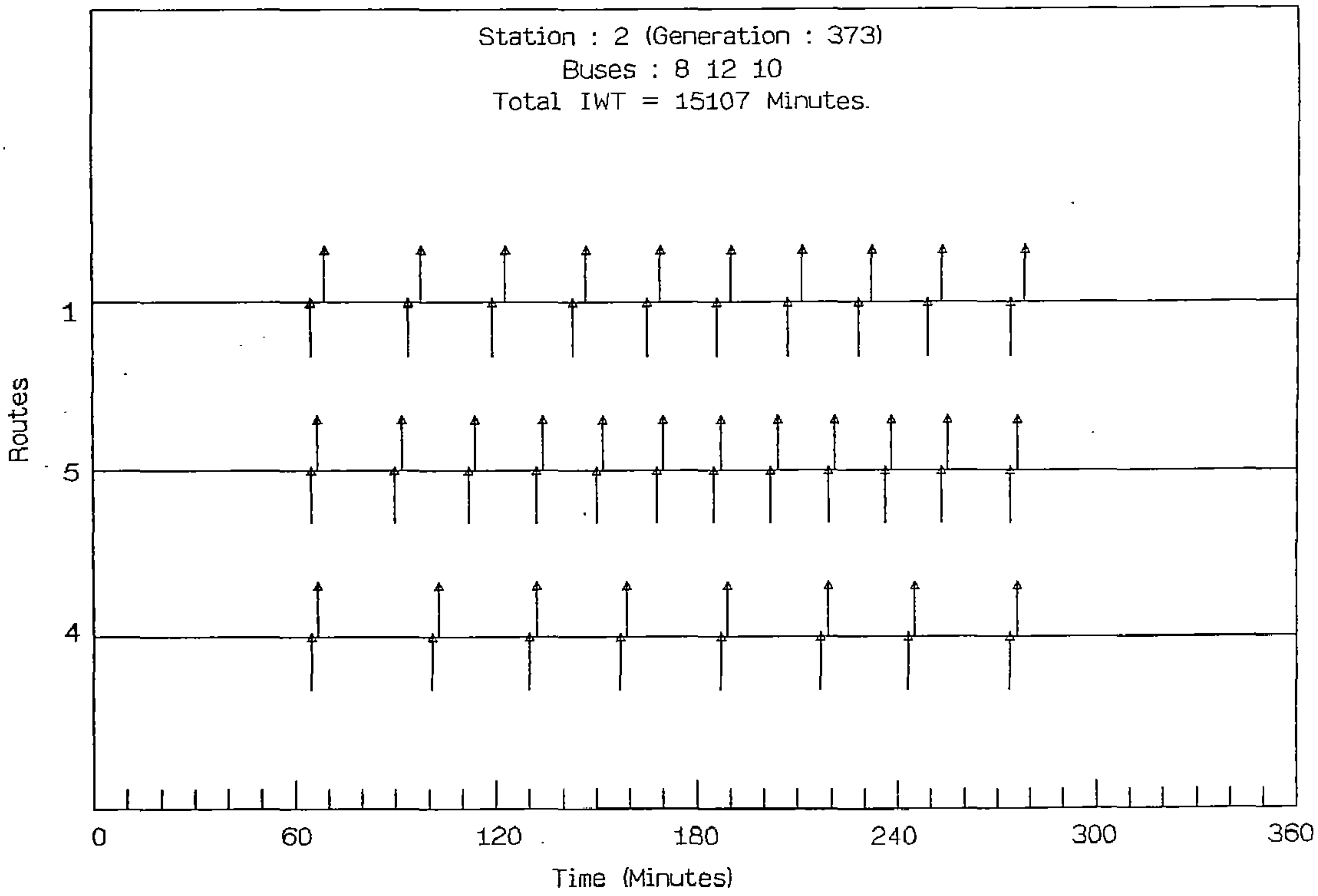


Figure 6.12(*b*): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

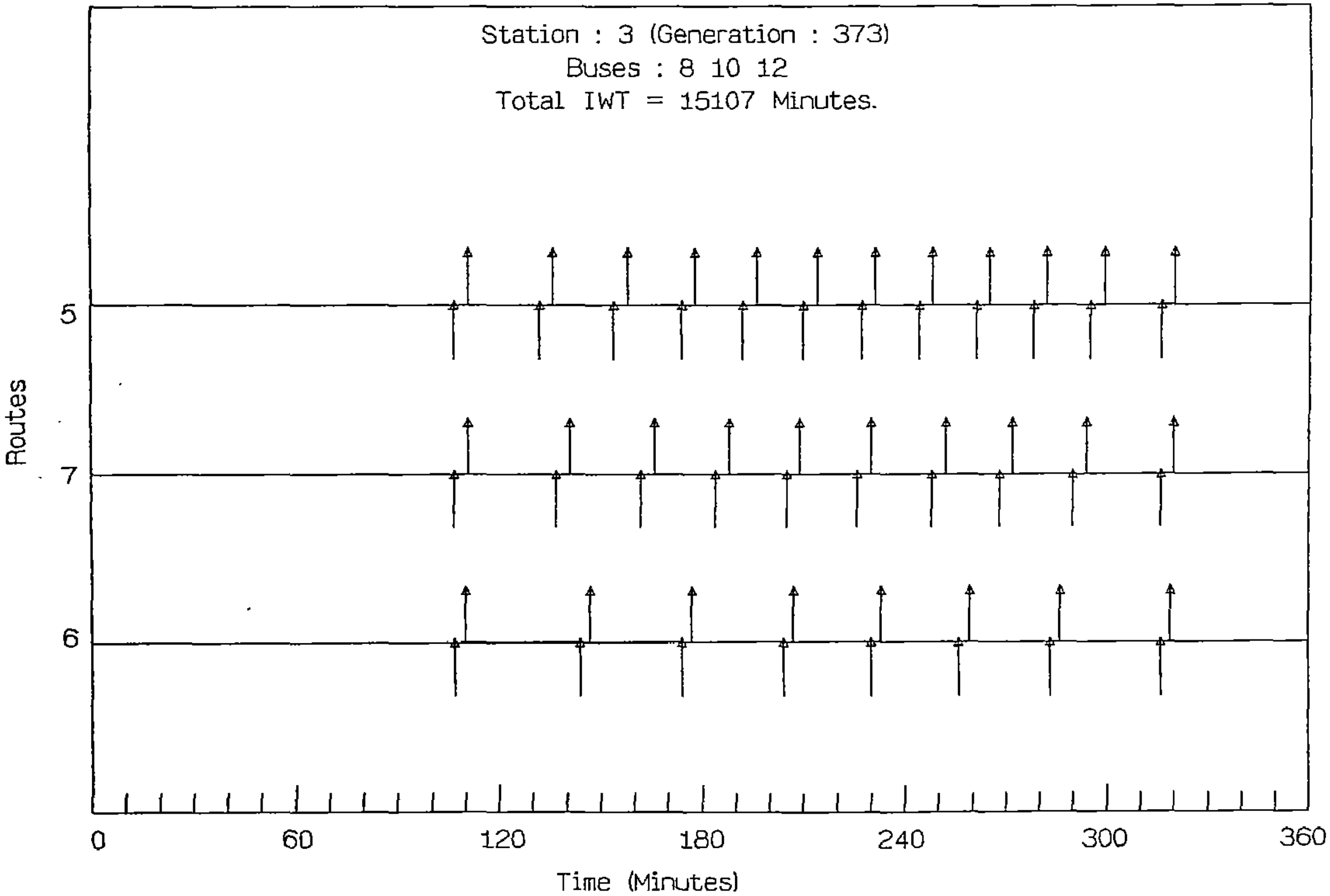


Figure 6.12(c): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

time to be minimum. For this case, the total initial waiting time obtained is 15107 minutes and as seen from the schedule, more buses with less headway come in the second half and the headways are more on the other half as the number of passengers arrive is less.

Case - V

Again in this case also as in case - III above, we cannot predict any intuitive optimal schedule due to the introduction of TT term into the objective function. The schedule obtained for TWT consideration with non-uniform arrival pattern of passengers is presented in the Figures 6.13(a), 6.13(b) and 6.13(c). The total waiting time obtained is 16540 minutes.

6.3.3.2 Optimal Schedules for Network - II

In this section, a generalized procedure for scheduling is developed. For the Network-I, the schedules were found at different stations at lagged scheduling windows of constant time range. A detailed description of the procedure was given in the previous section. However, this procedure cannot be applied in more complex networks such as Network-II because any station is connected to the others by two or more ways. The problem with applying this procedure (i.e. the procedure used for Network-I) is explained next.

In the Network-II shown in the Figure 2.3, there are routes connecting the Stations 1,2; 1,3 and 2,3. Thus for example, Station 1 is connected to Station 3 through Route 2 as well as via Station 2 (i.e. by Routes 1 and 5). Let the travel time on the Route 1 be a minutes, on Route 2 be b minutes and on Route 5 be c minutes and s_1, s_2 and s_3 be the stopping times of buses on the routes. According to the earlier procedure, since Stations 1 and 2 and Stations 2 and 3 are directly connected, the scheduling windows at Stations 2 and 3 should start at $a+s_1$ minutes (assuming scheduling window at Station 1 starts at 0 minutes) and $a+s_1+c+s_3$ minutes respectively. Again according to the earlier procedure, since Stations 1 and 3 are directly connected the scheduling window at Station 3 should start at $b+s_2$ minutes. However, the time $b+s_2$ minutes may be less than, greater than or equal to $a+s_1+c+s_3$. When these two times are equal, all the buses which starts within a 240 minutes schedule time window at a station will come

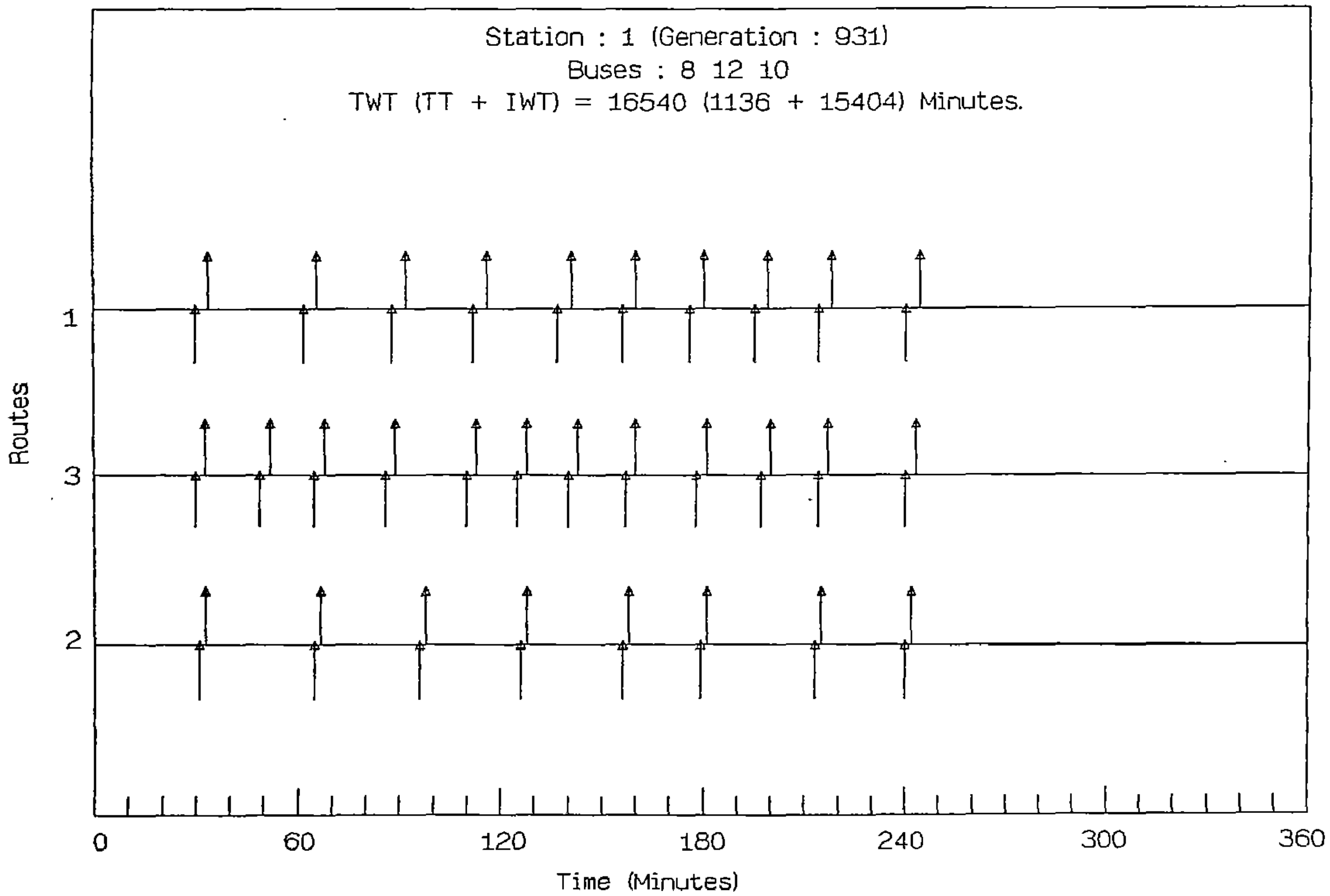


Figure 6.13(*a*): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

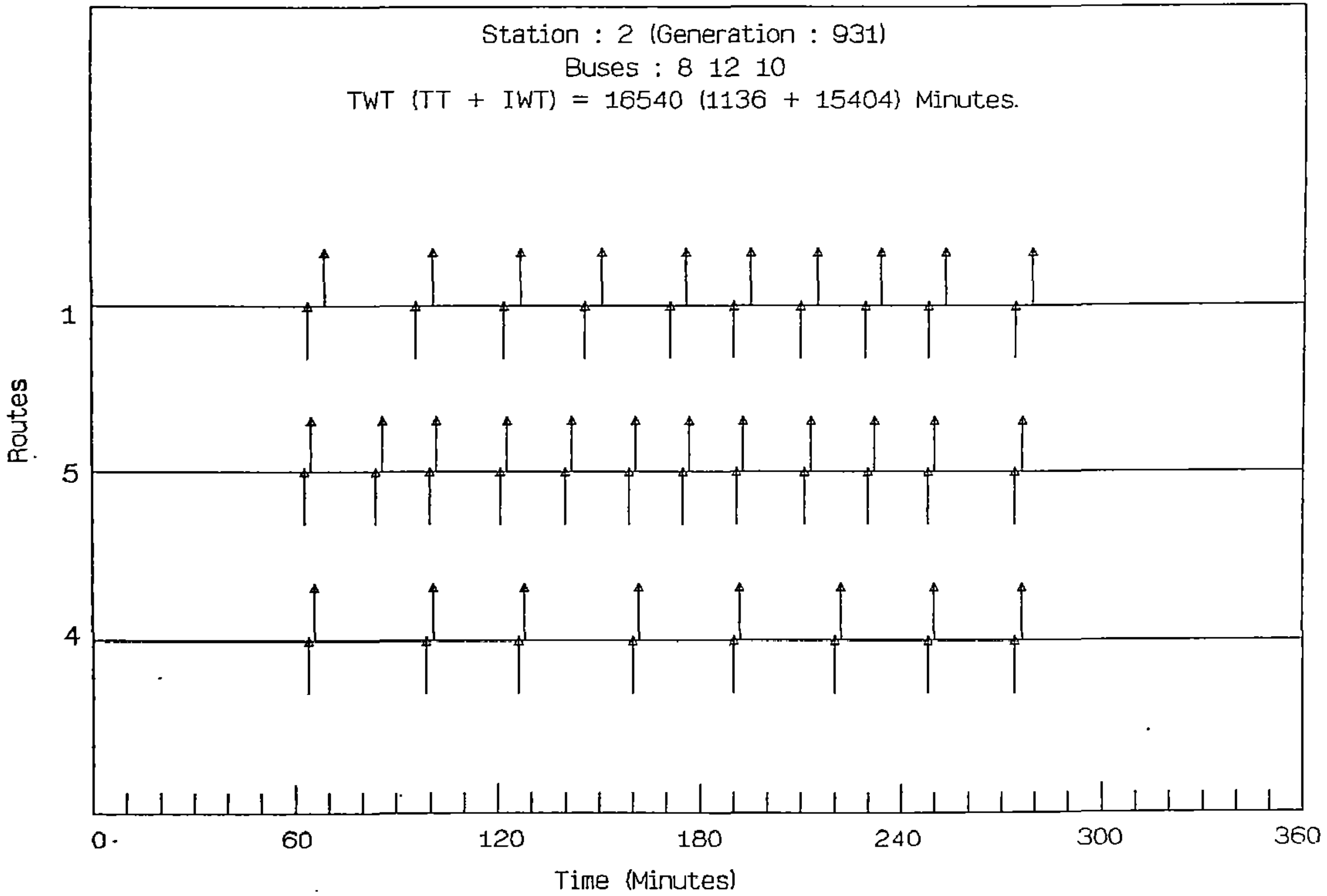


Figure 6.13(b): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

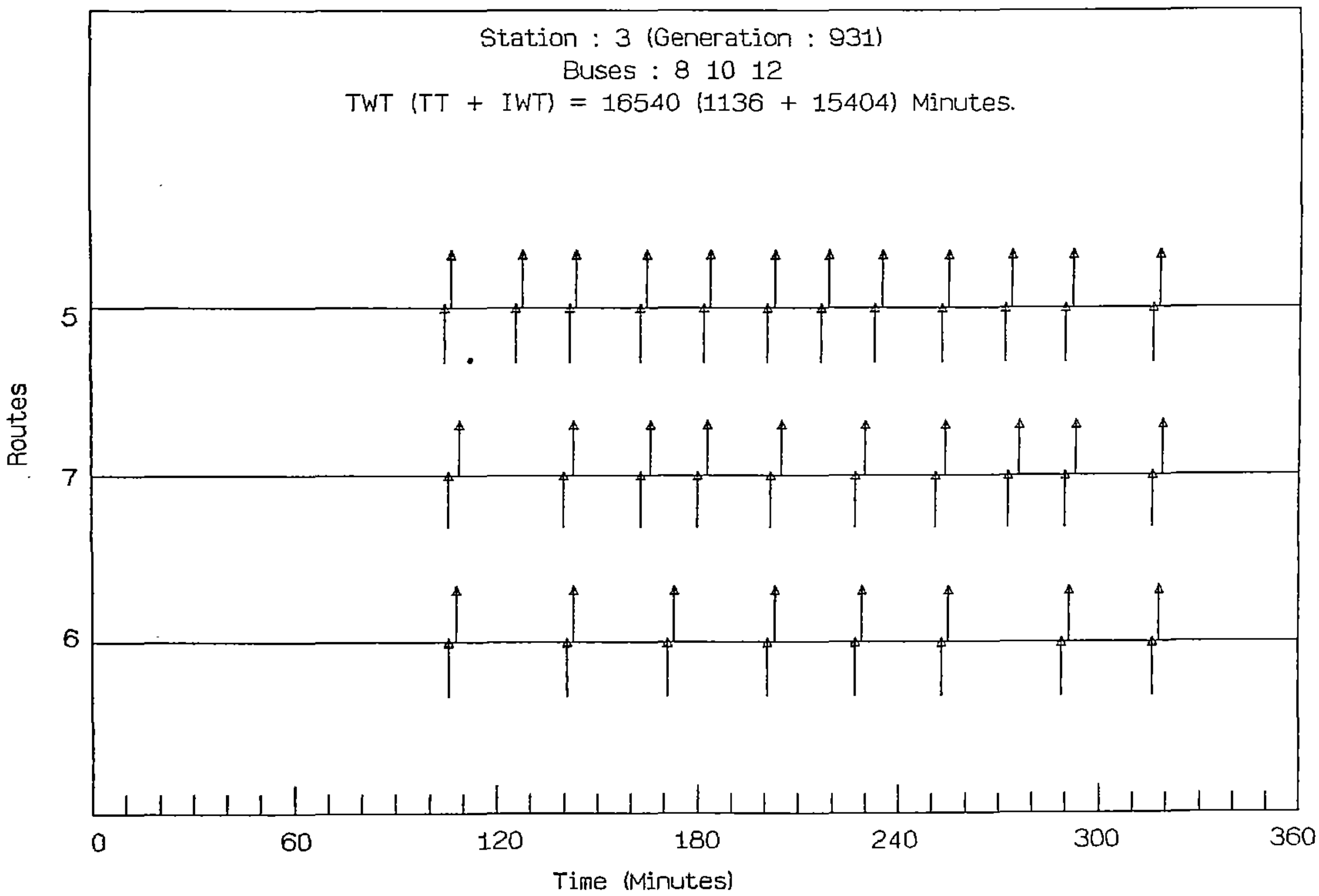


Figure 6.13(c): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

within another 240 minutes window at the following station shifted by the travel time plus the corresponding stopping time between the two stations. Then the procedure used for Network-I can be applied for this network also without any problem. However, we cannot expect $a + s_1 + c + s_3$ to be equal to $b + s_2$ for any general network. Thus, in the general case there will exist an ambiguity as to when the scheduling window should start (say, at Station 3). Similarly a conflict will exist as to when the scheduling period should end. This conflict, however, is not something inherent to the system. It is artificially created by the earlier procedure. The fact is, the schedule repeats itself on either side of any scheduling period. And this is something which is continuous rather than bound. The revised and more general procedure proposed here acknowledges and uses this fact to achieve a schedule.

Consider three equal scheduling time windows at a station in a transit system network. Let there be two routes P and Q at the station and the optimal schedule of buses on these routes is as shown in the Figure 6.14 for a particular arrival pattern of passengers. Let the time windows for Routes P and Q be pq, qr, rs and $p'q'$, $q'r'$, $r's'$ respectively. Note that there is a lag in the scheduling periods for the two routes. It can be considered that buses on the Route Q come from another station with that lag.

Let there are 4 buses on Route P and 5 on Route Q and these are numbered as shown. Now, as earlier, if we were to consider the periods qr and $q'r'$ only and calculate the transfer times based on these periods, then we

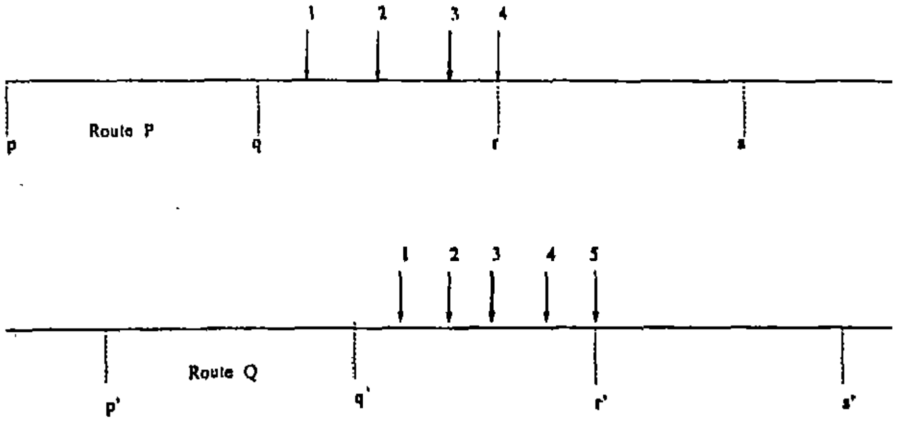


Figure 6.14: Schedule in one scheduling period

could see that the transfers from all buses on Route P is possible to Route Q. But the transfers from buses 4 and 5 on Route Q are not possible to Route P, as there are no buses after time r on Route P. So by the previous procedure, the transfer times are undefined for the passengers transferring from buses 4 and 5 on Route Q. However this neither happens nor is acceptable.

To avoid the above conflict and to make the schedule continuous over a period of time, the schedule in the periods qr and $q'r'$ on the routes P, Q are repeated on either side of the scheduling periods (as one of the assumptions is that the arrival pattern of passengers is repeated with an interval of pq over time). Now the schedule is as shown in the Figure 6.15. Note that only the headways between solid arrows are decision variables in the optimization problem and the arrivals with broken arrows are merely repetitions of the arrivals of corresponding buses during the period qr. Now, the transfer times for buses 4 and 5 on Route Q are no longer undefined. It is noted, however, that the transfer from, say the first bus (with broken arrow) of Route Q to the third bus (with broken arrow) of Route P is not considered since the same has already been considered by calculating the transfer time from the first bus (with solid arrow) of Route Q to the third bus (with solid arrow) of Route P.

The scheduling period pq in the present study is 240 minutes. The above procedure can be applied to the Network-II as follows. In the Network-II, there are three routes at the Station 1, out of which one is isolated and the other two go to the Stations 2 and 3. On all these routes, buses are scheduled for 0 to 240 minutes, and this schedule is repeated after 240 minutes. However, there may be routes at a station, where buses will arrive from other

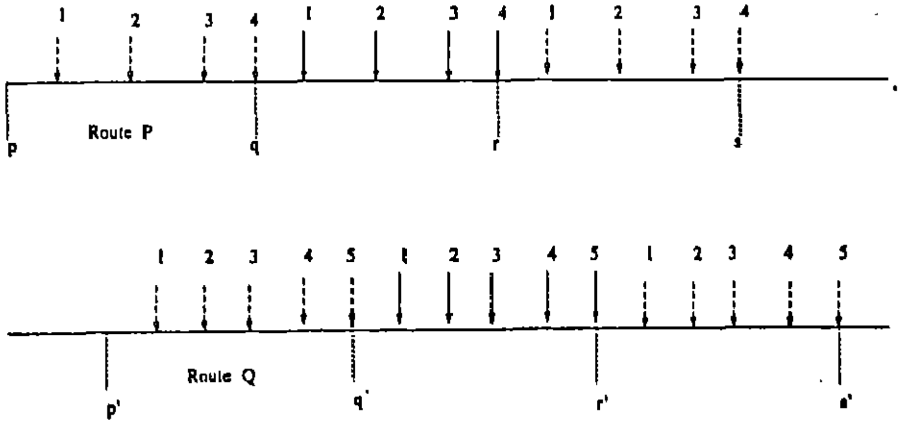


Figure 6.15: Schedule on either of a scheduling period

stations with a lag of travel time plus their stopping time at the first station. As on Route 1 in the Network-II, the buses will come at $30+s_1$ minutes upto $240 + 30 + s_1$ minutes at Station 2, since these buses started from Station 1 between 0 to 240 minutes. From 0 to $30 + s_1$ minutes at Station 2, the headways of buses on Route 1 are the same as in interval 240 minutes to $240+30+s_1$ minutes (that is the last $30+s_1$ minutes of the scheduling period $30+s_1$ to $240+30+s_1$ minutes). After $240+30+s_1$ minutes, the headways of buses on Route 1 are the same as in the interval $30+s_1$ to $240+30s_1$ minutes period. However on the Routes 4 and 5 at Station 2, buses are scheduled from 0 to 240 minutes, and the same headways repeated after 240 minutes on these routes. Similar is the case at Station 3 in Network-II.

In this procedure, the transfer time and initial waiting time for passengers are calculated based on buses arriving within a period of 240 minutes (since 240 minutes is the scheduling window), though the schedule is shown for a period greater than 240 minutes at each of the stations. For the present network, the transfers between the Routes 1 and 5 at the Station 2 are not considered as there is a connecting route between the Stations 1 and 3 and hence one would not transfer from Route 1 to Route 5 at Station 2. The same uniform and non-uniform arrival patterns used for Network-I are used for this network also.

In the following pages, the results obtained for the Network-II are discussed. The results from the new procedure as expected, follows the pattern of the results from the previous network. The following discussion of the results may thus seem repitation but is provided here to show the applicability of the modified procedure presented here.

Equal resource on each route

In the following pages, the optimal schedules obtained for various cases defined earlier for equal resource on each route are presented.

Case - I

The optimal schedule based on only transfer time consideration with equal resource on each route is presented in the Figures 6.16(a), 6.16(b) and 6.16(c). In the figures, the arrival and departure time for various buses are represented

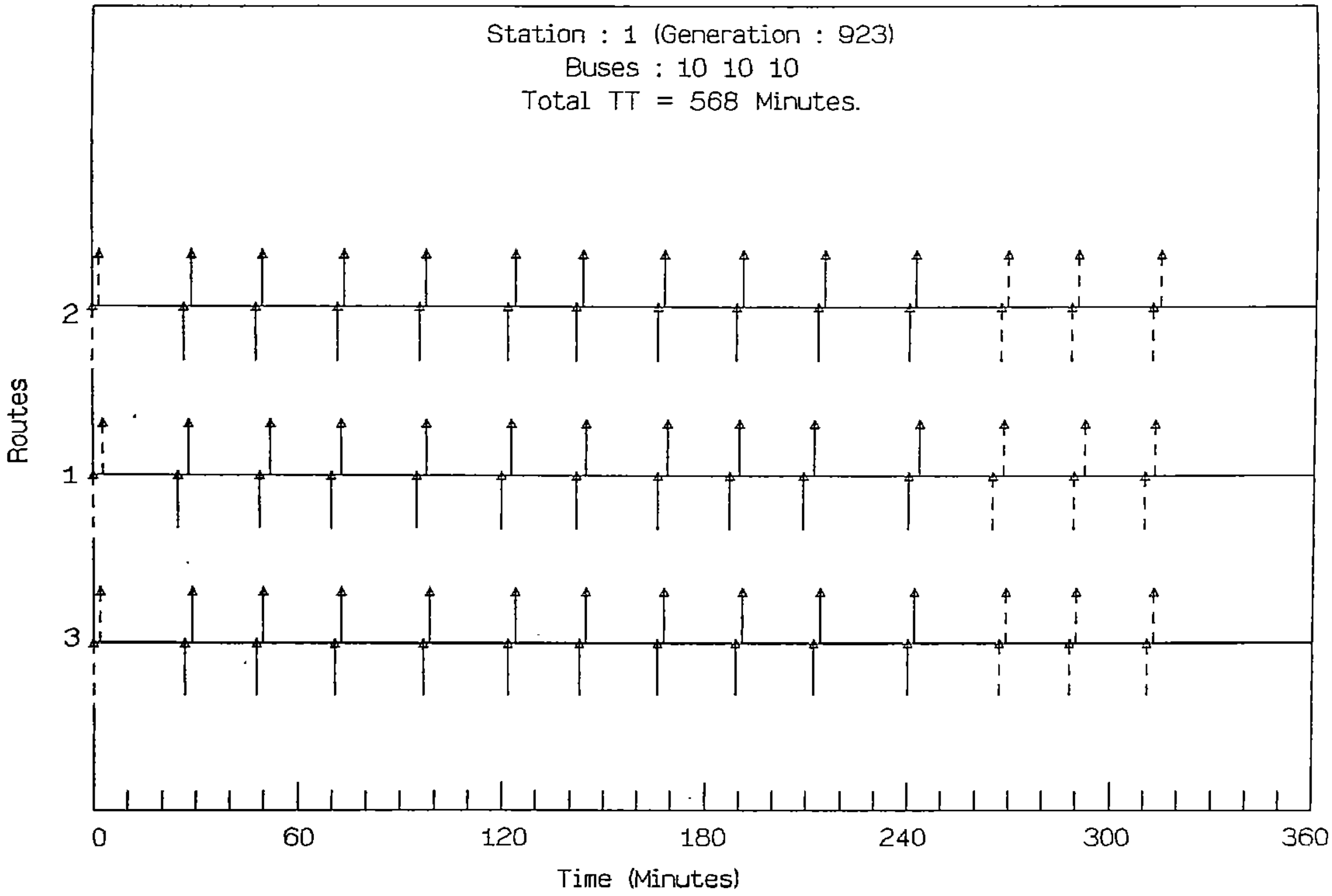


Figure 616(a): Optimal schedule for TT consideration

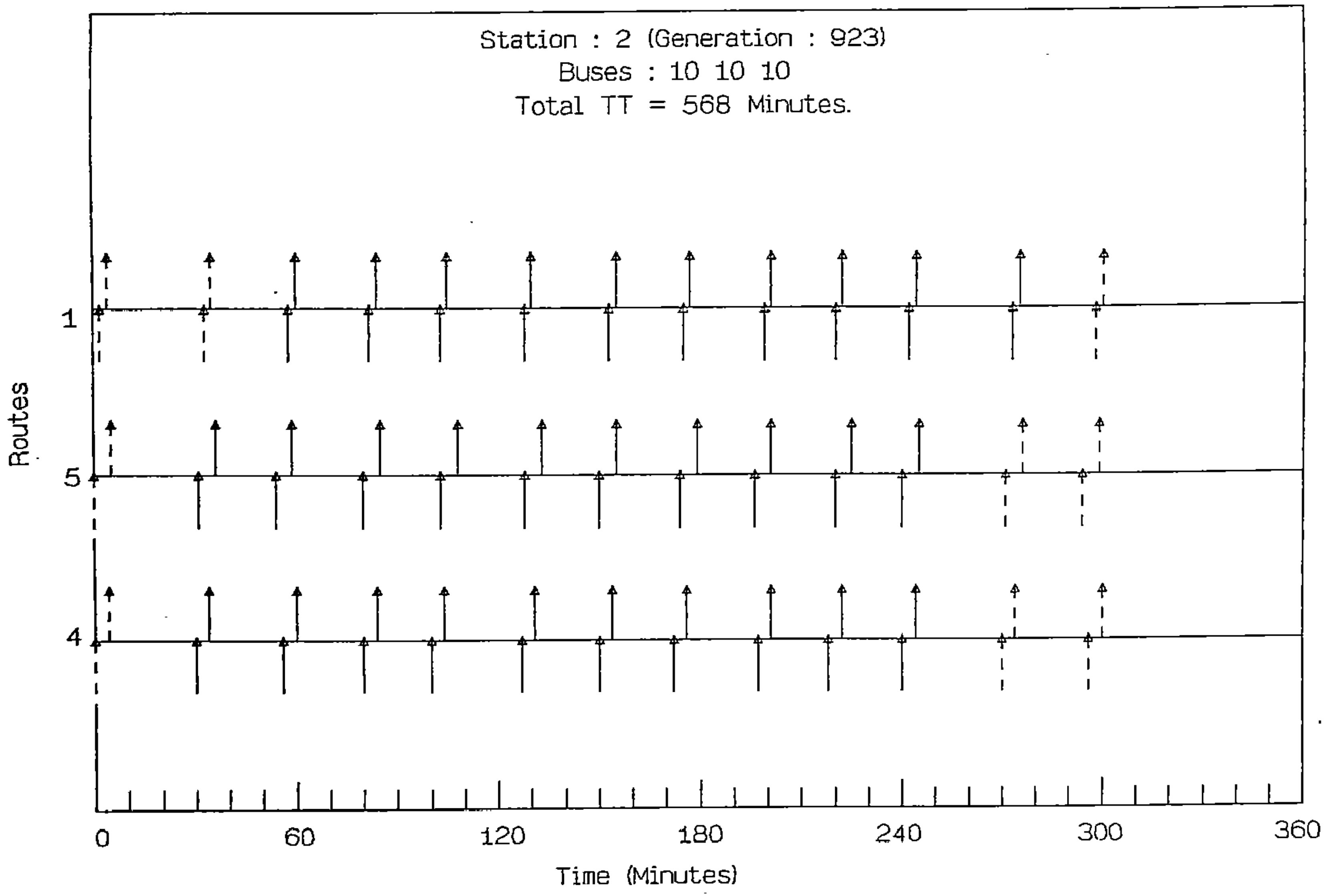


Figure 6.16(*b*): Optimal schedule for TT consideration

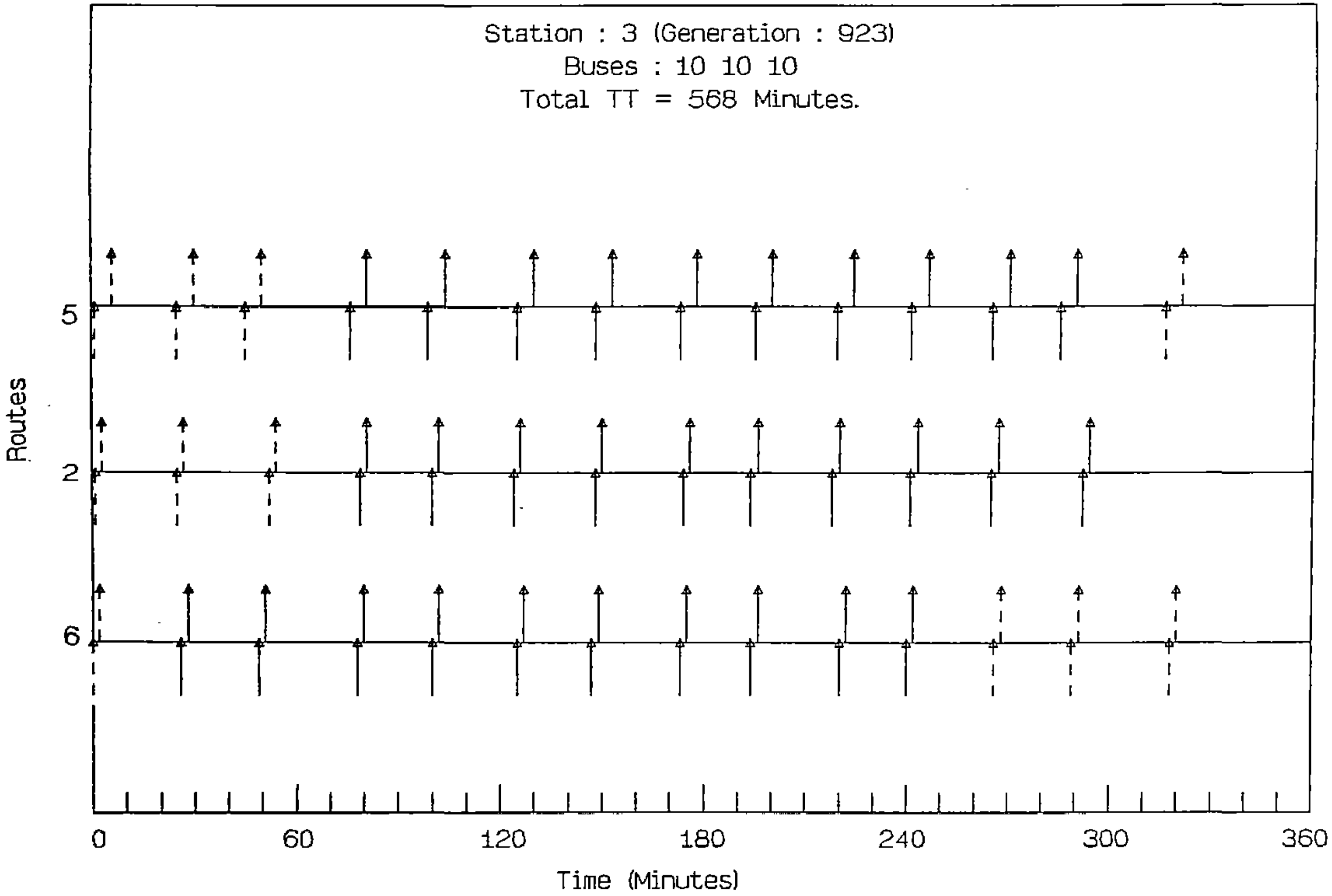


Figure 6.16(c): Optimal schedule for TT consideration

by the solid arrows as well as the broken arrows. The solid arrows represent the actual decision variables present in the optimization problem. The broken arrows are fixed based on the above variables. At the Station 1, the solid arrows are present from 0 to 240 minutes on all the routes as buses start at this station. The bus at 0 minutes may be thought of as the last bus of the previous scheduling period. Similarly the first bus after 240 minutes is the same as the first bus after 0 minutes. The arrival and departure times of various buses at each of the stations are represented upto to a time of about 240 minutes plus maximum of all the travel times. At the Station 2, the solid arrows are present from 0 to 240 minutes on Routes 4 and 5 and around 30 to 270 minutes on the Route 1. On the Route 1, buses come from Station 1. On this route, the buses within 0 to 30 minutes, are same as the penultimate bus and the last bus in the 30 to 270 minutes. After 270 minutes, the headways within 30 to 270 minutes are repeated. At the Station 3 also the same procedure is adopted.

The schedule shown in the figures are expected as the buses coming on various routes at a station at more or less same time is the optimal schedule for only transfer time consideration with equal resource on each route. This schedule is independent of arrival pattern of passengers and is one of the infinite number of optimal solutions for this case. The total transfer time obtained is 568 minutes.

Case - II

The schedule obtained for IWT consideration with uniform arrival pattern of passengers is shown in the Figures 6.17(a), 6.17(b) and 6.17(c). The total IWT obtained for this case is 15570 minutes. As seen from the figures, the headways between any two successive buses on a particular route are same which is the optimal solution for IWT case with uniform arrival pattern of passengers.

Case - III

In this case, the optimal schedule for TWT consideration with uniform arrival pattern of passengers is derived. The schedule is shown in the Figures 6.18(a), 6.18(b) and 6.18(c). This schedule is also good as buses arrive at

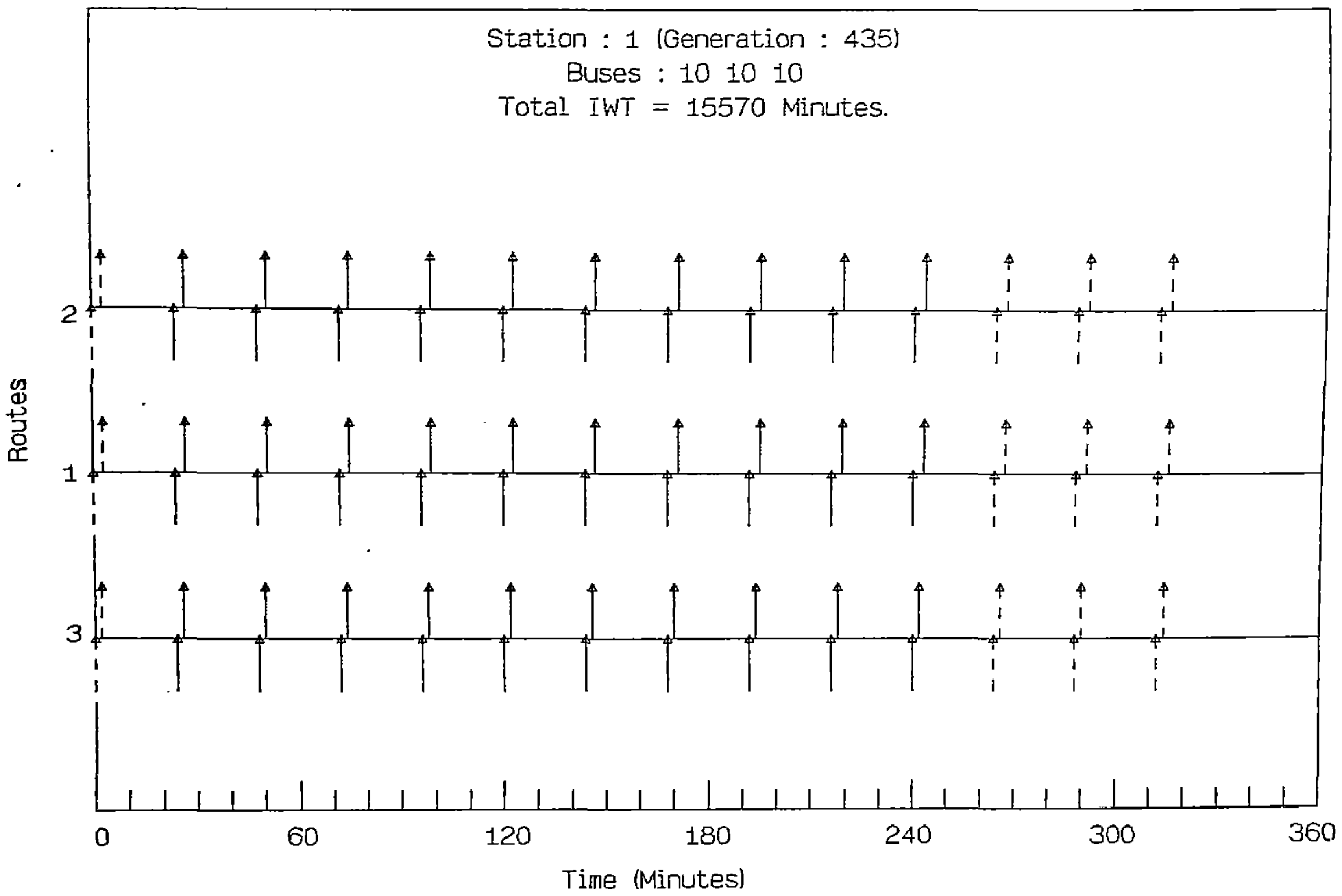


Figure 6.17(a): Optimal schedule for IWT consideration (Uniform arrival pattern)

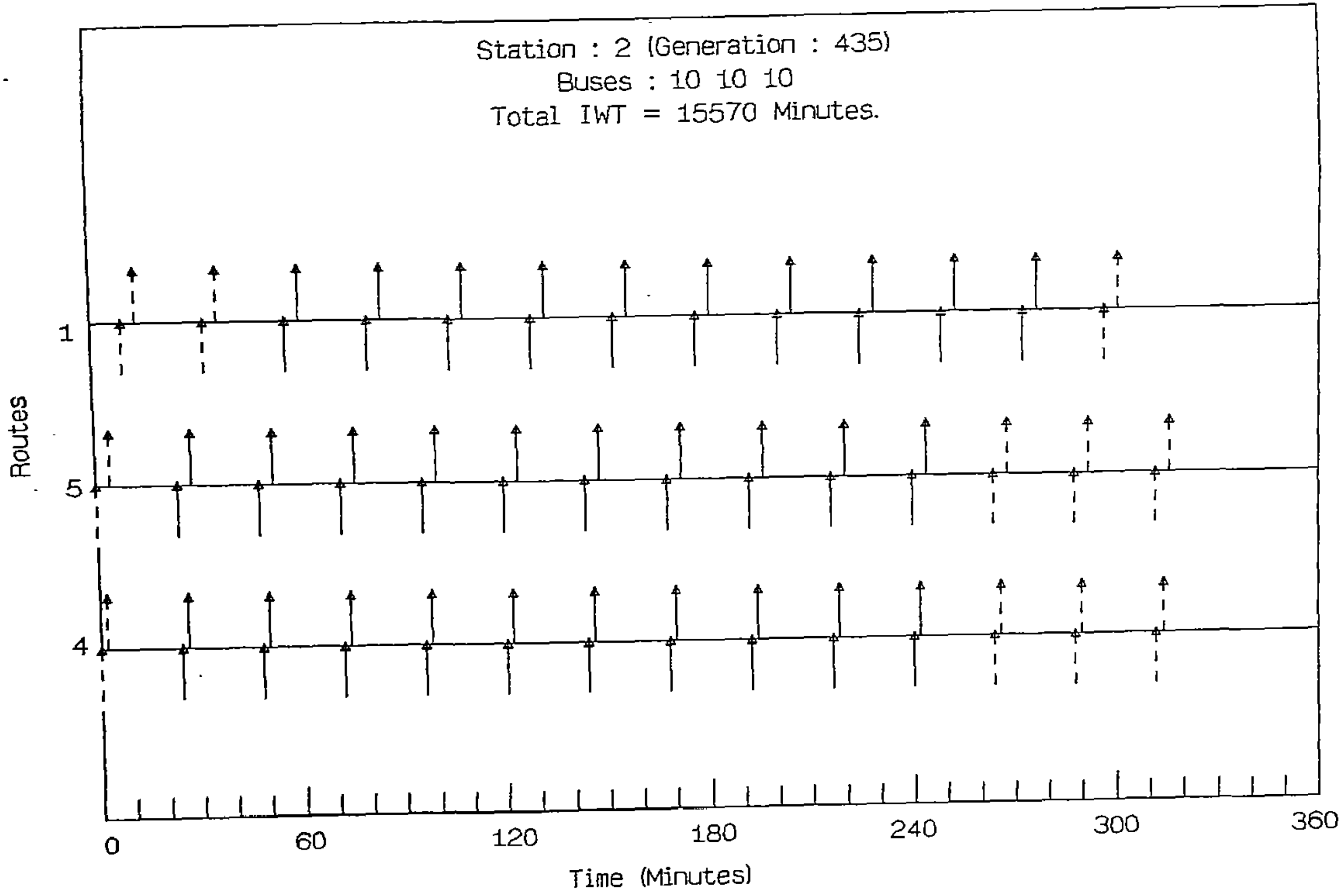


Figure 6.17(b): Optimal schedule for IWT consideration (Uniform arrival pattern)

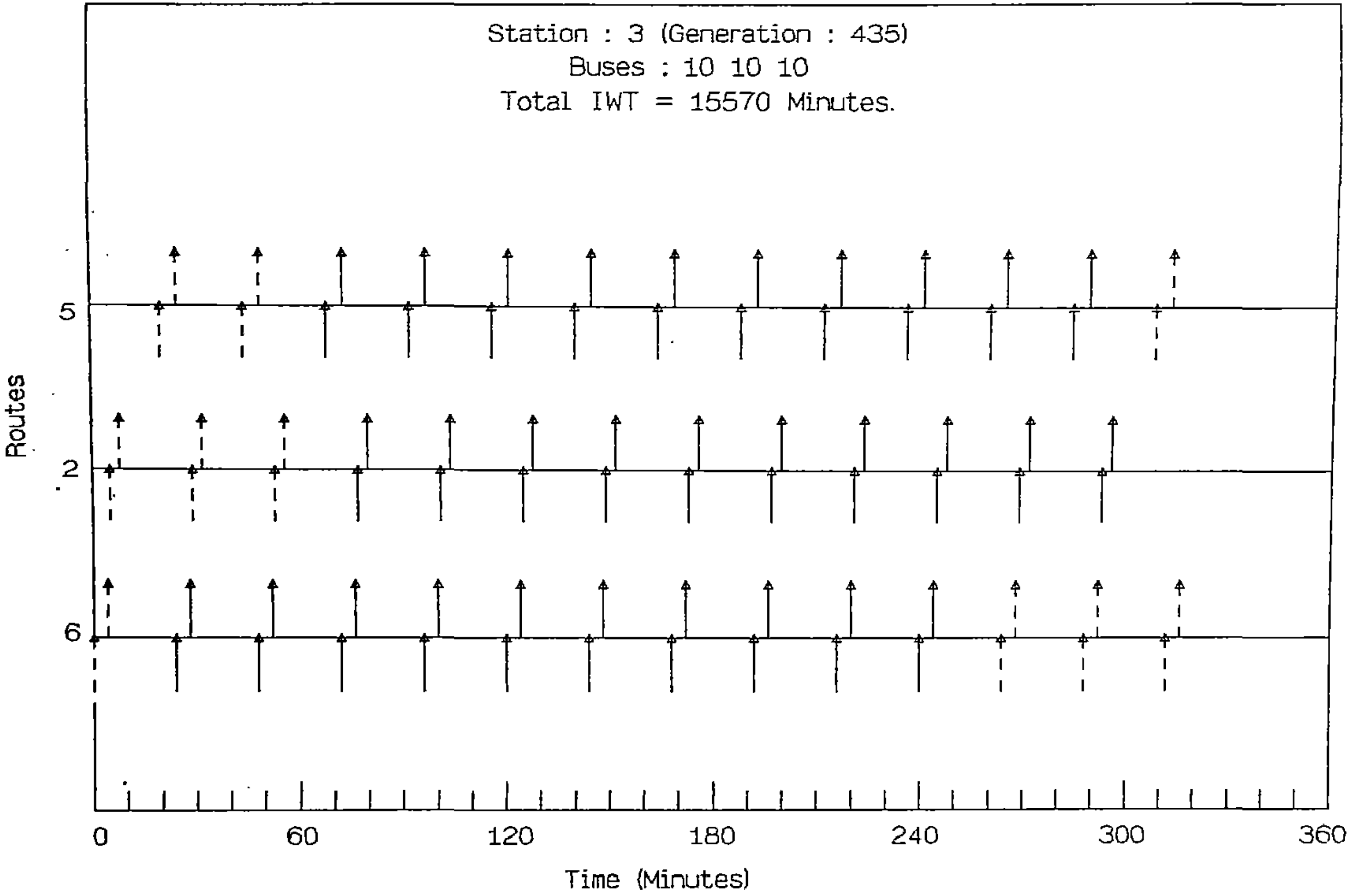


Figure 6.17(c): Optimal schedule for IWT consideration (Uniform arrival pattern)

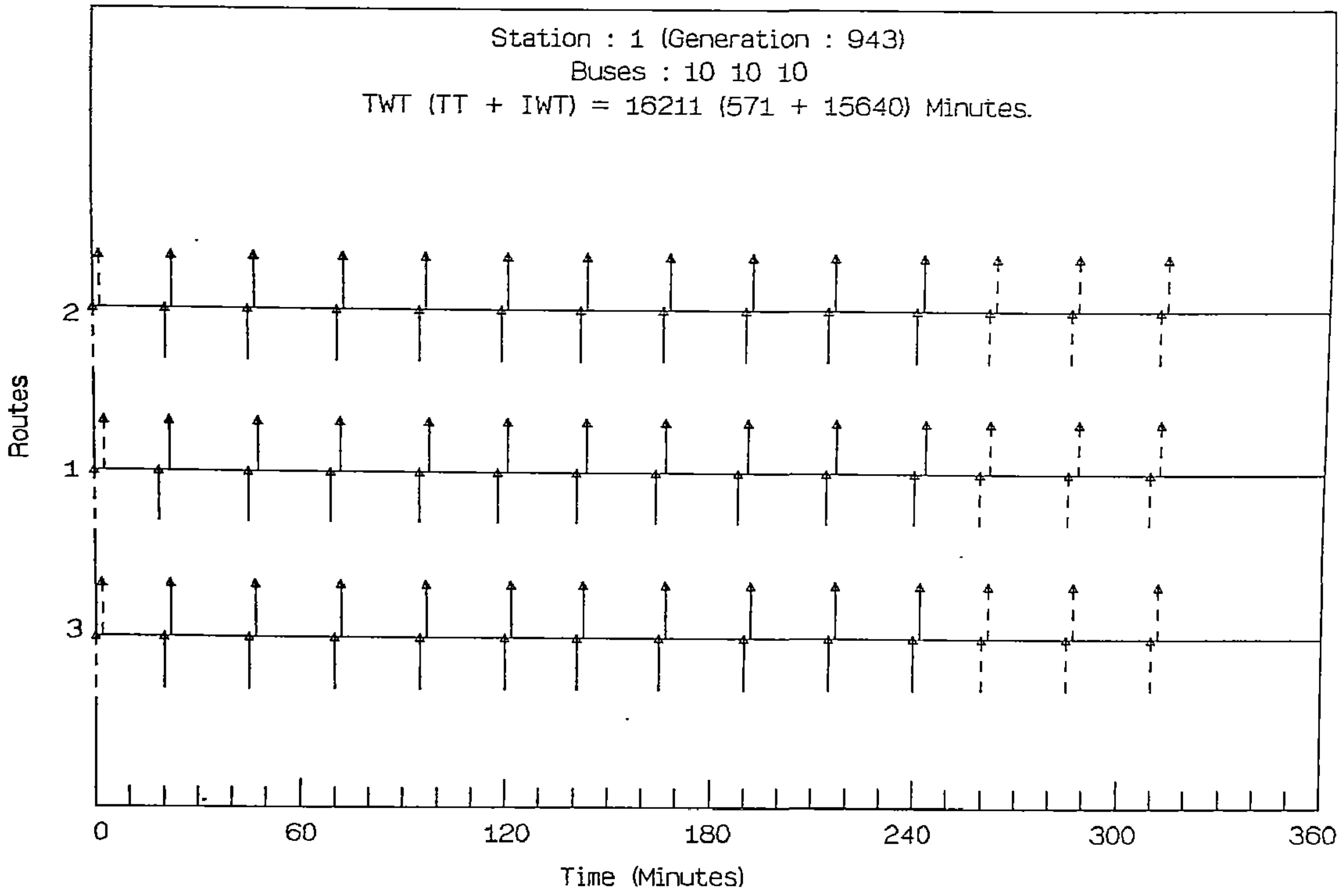


Figure 6.18(a): Optimal schedule for TWT consideration (Uniform arrival pattern)

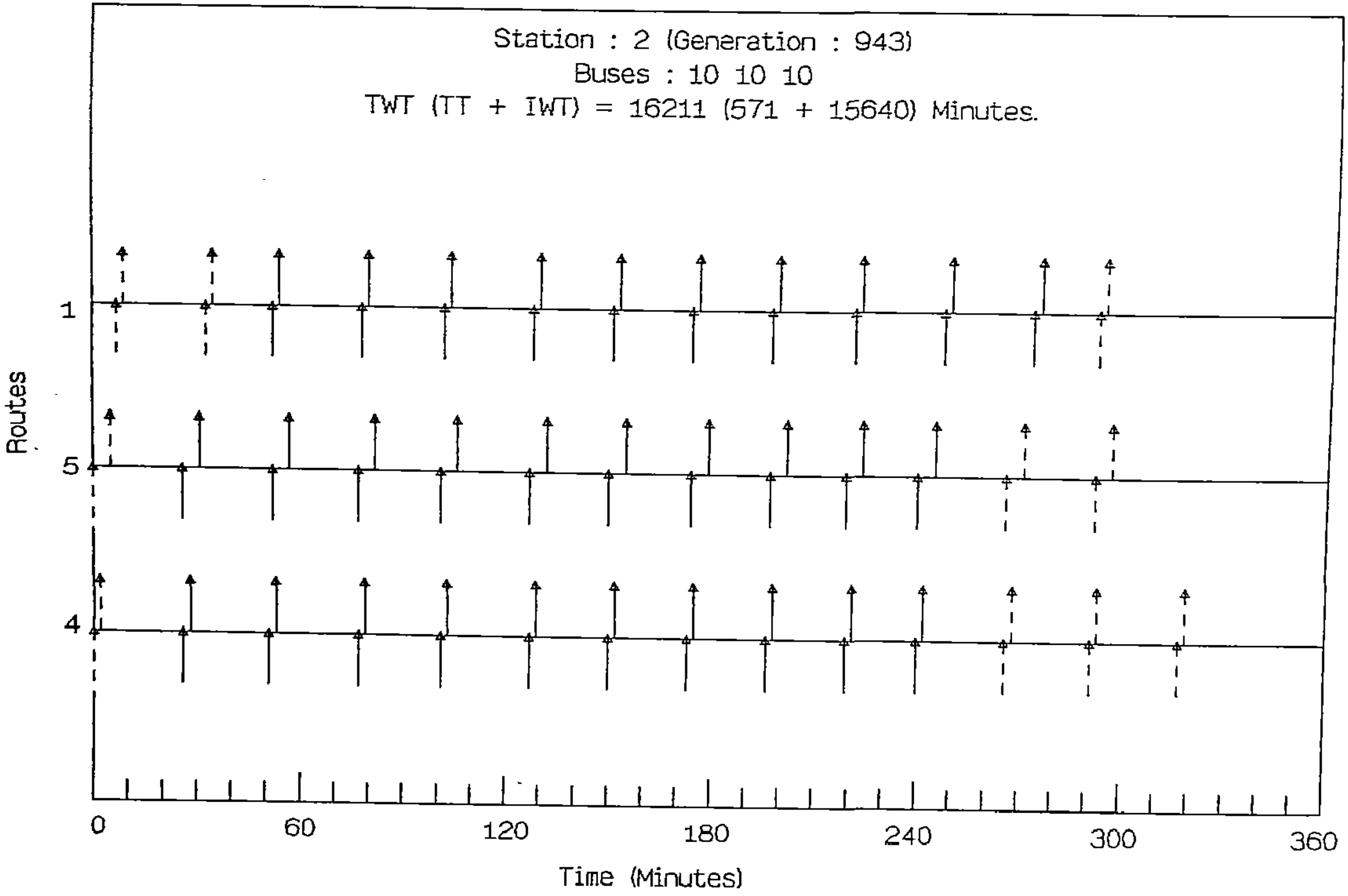


Figure 6.18(*b*): Optimal schedule for TWT consideration (Uniform arrival pattern)

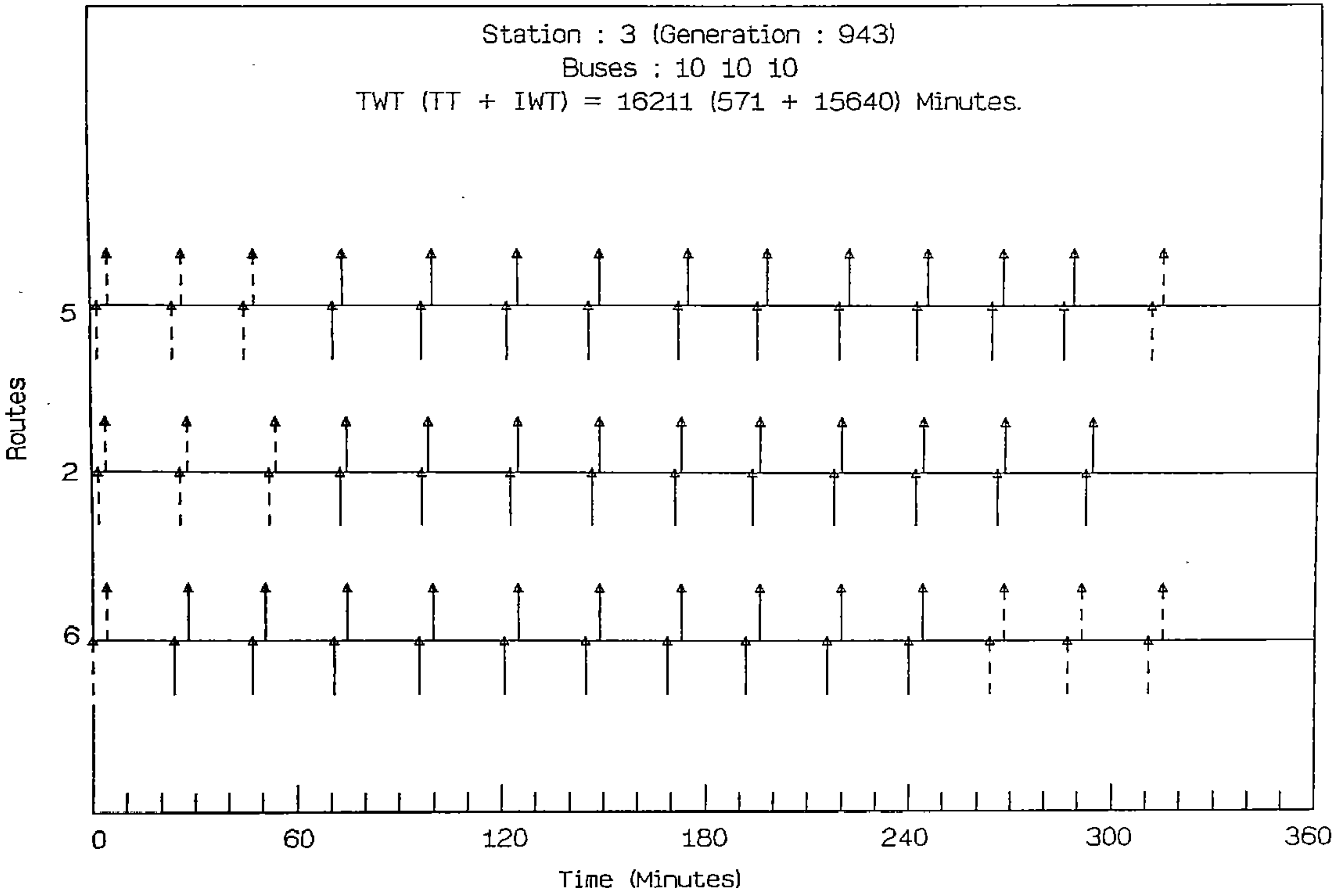


Figure 6.18(c): Optimal schedule for TWT consideration (Uniform arrival pattern)

uniform headway on different routes at a station and the arrival times of corresponding buses on different routes at a station are more or less same. But on Route 1 at Station 2 and on Route 5 at Station 3, the arrival times of some buses (with broken arrows) are slightly different from that on other routes as they come with a lag with from other stations. The total waiting time obtained in this case is 16211 minutes.

Case - IV

The schedule for IWT case with non-uniform arrival pattern of passengers is shown in the Figures 6.19(a), 6.19(b) and 6.19(c). As the envelope function which defines the peaks of the arrival pattern of passengers has maximum height at 180 minutes (where more passengers come), more buses should come at and around that time and their headways should be less. As closely seen the figures, we can observe the same. The total IWT obtained for this case is 15178 minutes.

Case - V

The optimal schedule for TWT consideration with non-uniform arrival pattern of passengers is shown in the Figures 6.20(a), 6.20(b) and 6.20(c). More buses should come with less headways where the number of passengers coming is more. As the peak of the envelope function is at 180 minutes, more buses buses come between 120 to 240 minutes when compared to that in 0 to 120 minutes which can be understood from the figures. The TWT obtained is 15926 minutes.

Unequal resources on each route

In the above five cases, equal number of buses (i.e. 10) was considered on all the routes. In this section, the optimal schedules obtained with unequal resource on each route are presented for the five cases.

Case - I

The optimal schedule for only TT consideration with unequal resource on each route is shown in the Figures 6.21(a), 6.21(b) and 6.21(c). The total transfer time obtained for this case is 933 minutes, which is more than the

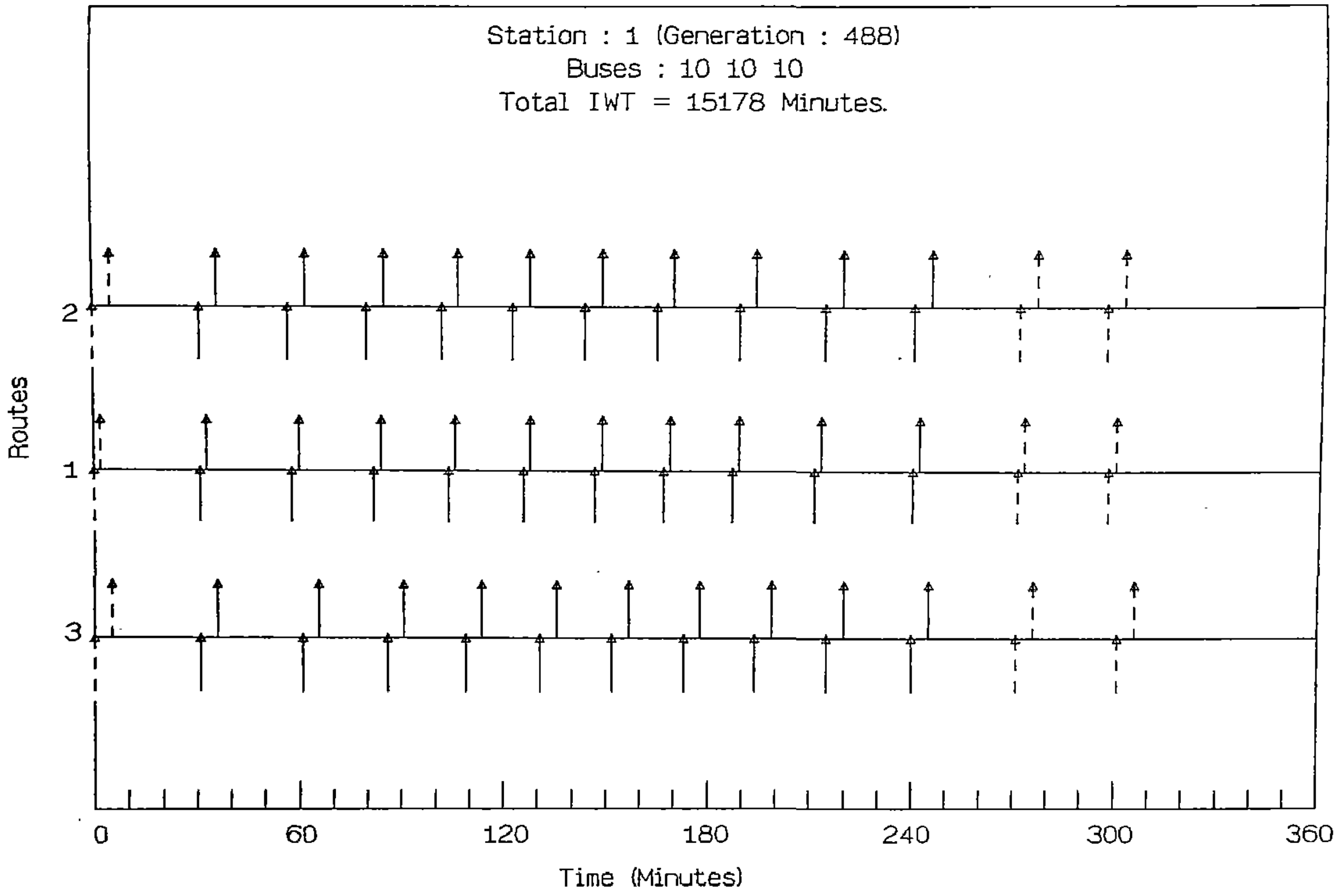


Figure 6.19(a): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

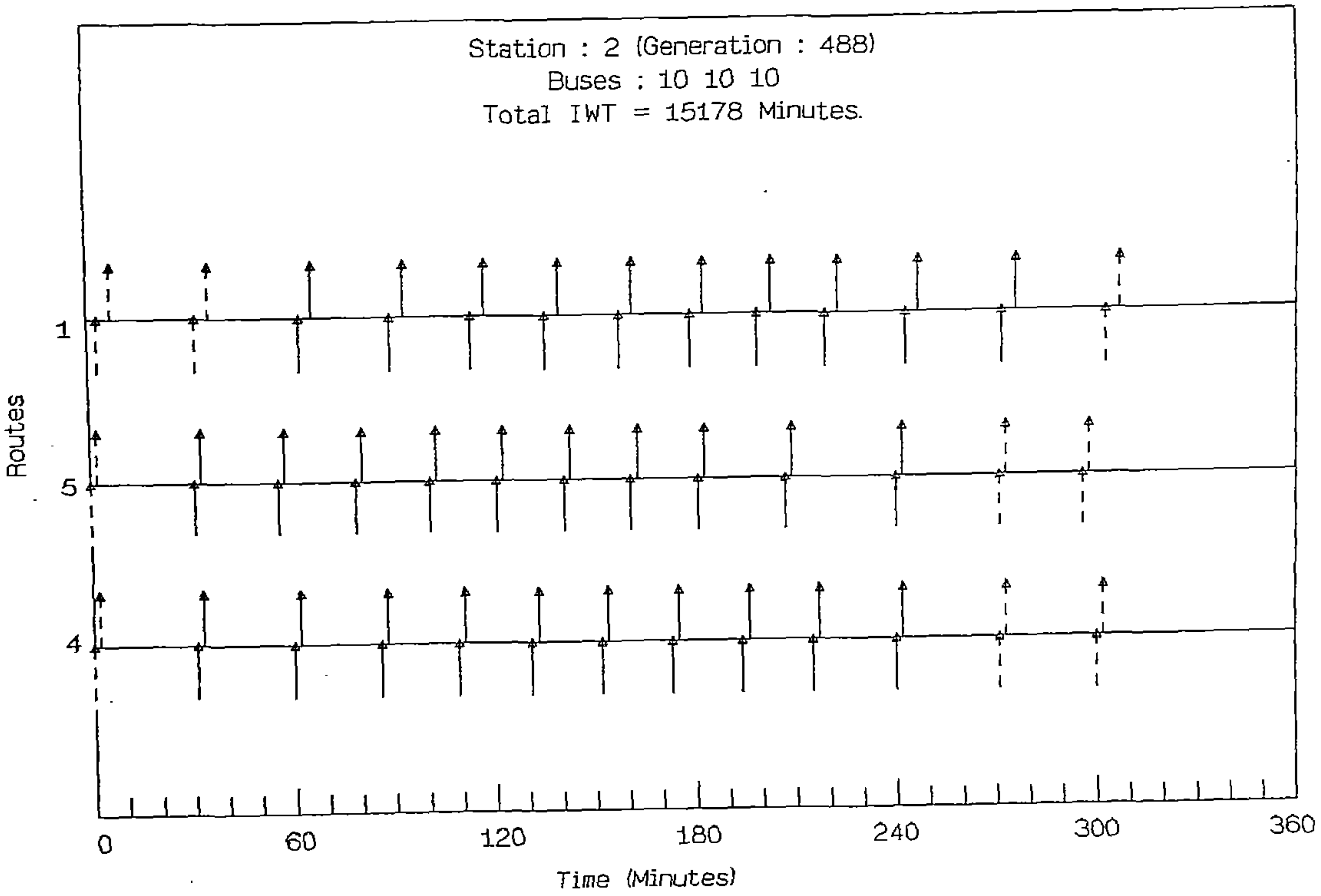


Figure 6.19(b): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

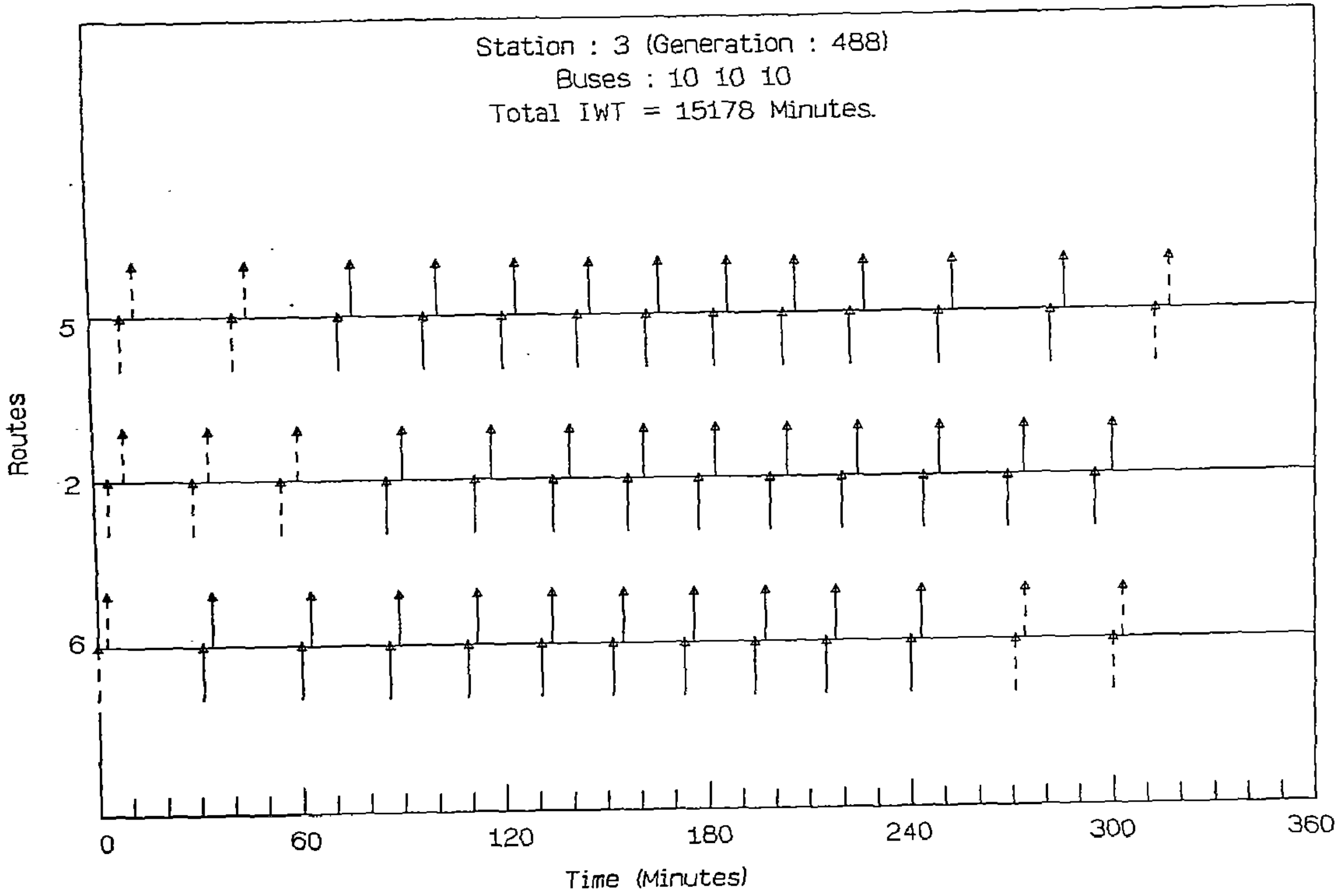


Figure 6.19(c): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

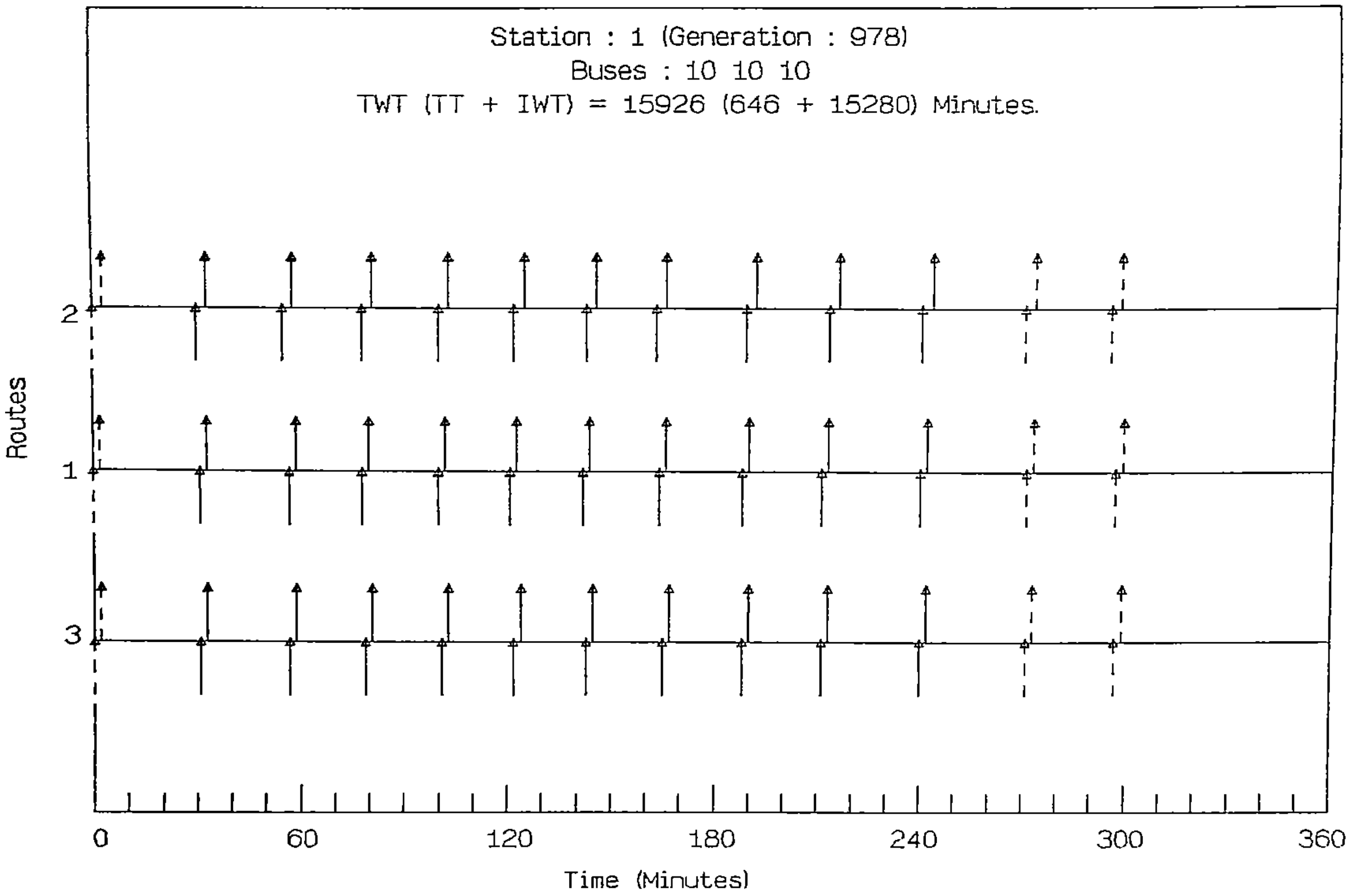


Figure 6.20(*a*): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

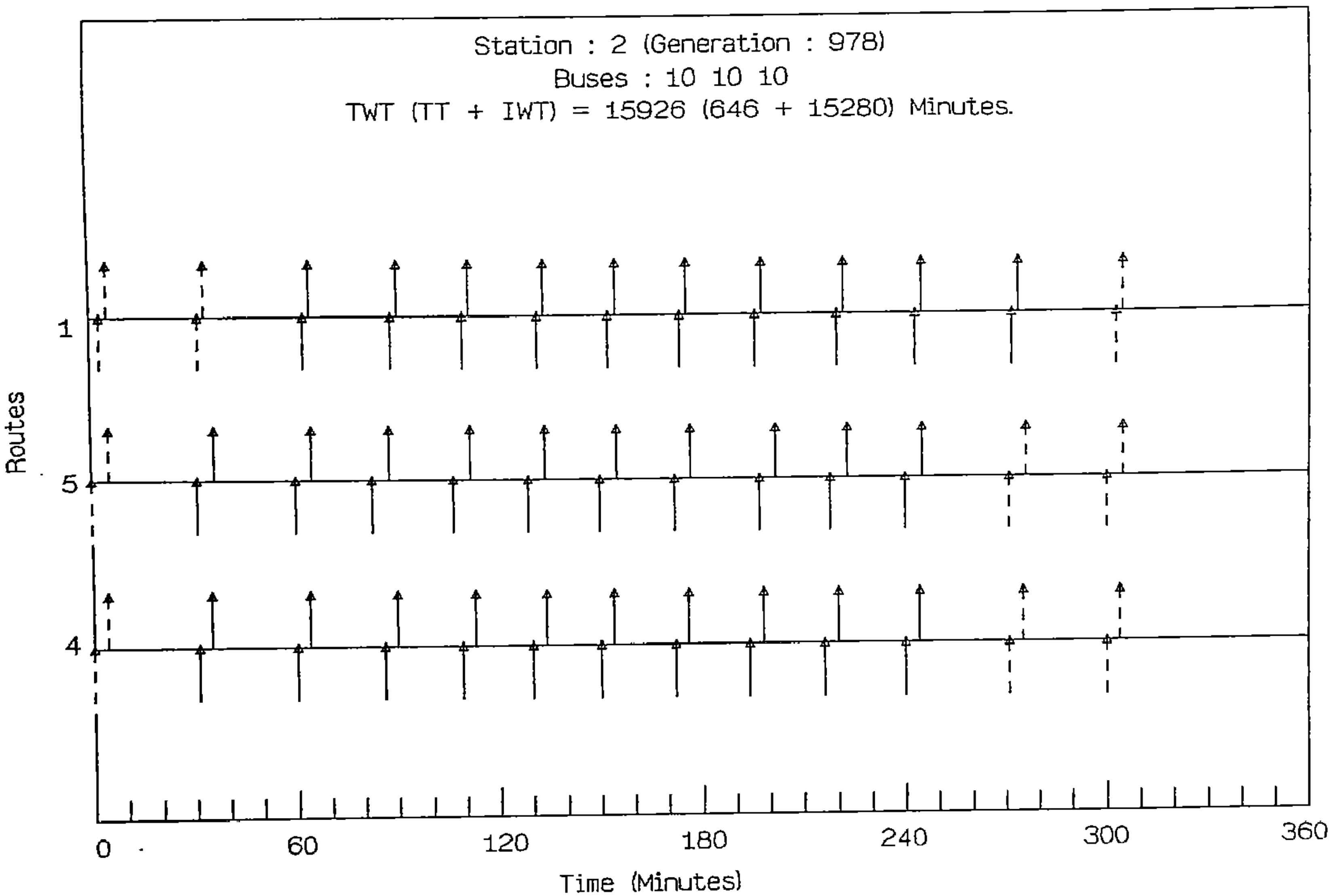


Figure 6.20(b): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

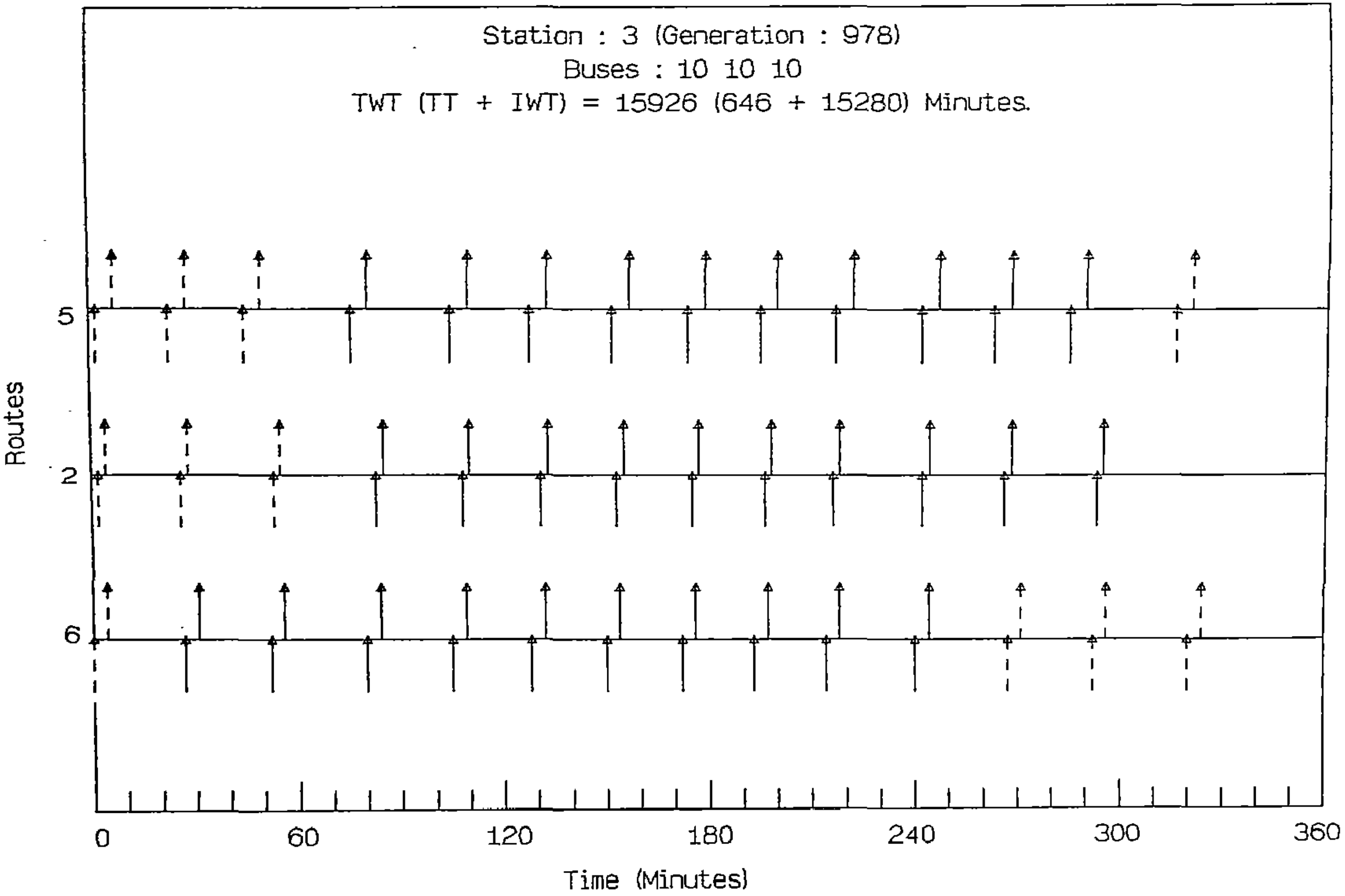


Figure 6.20(c): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

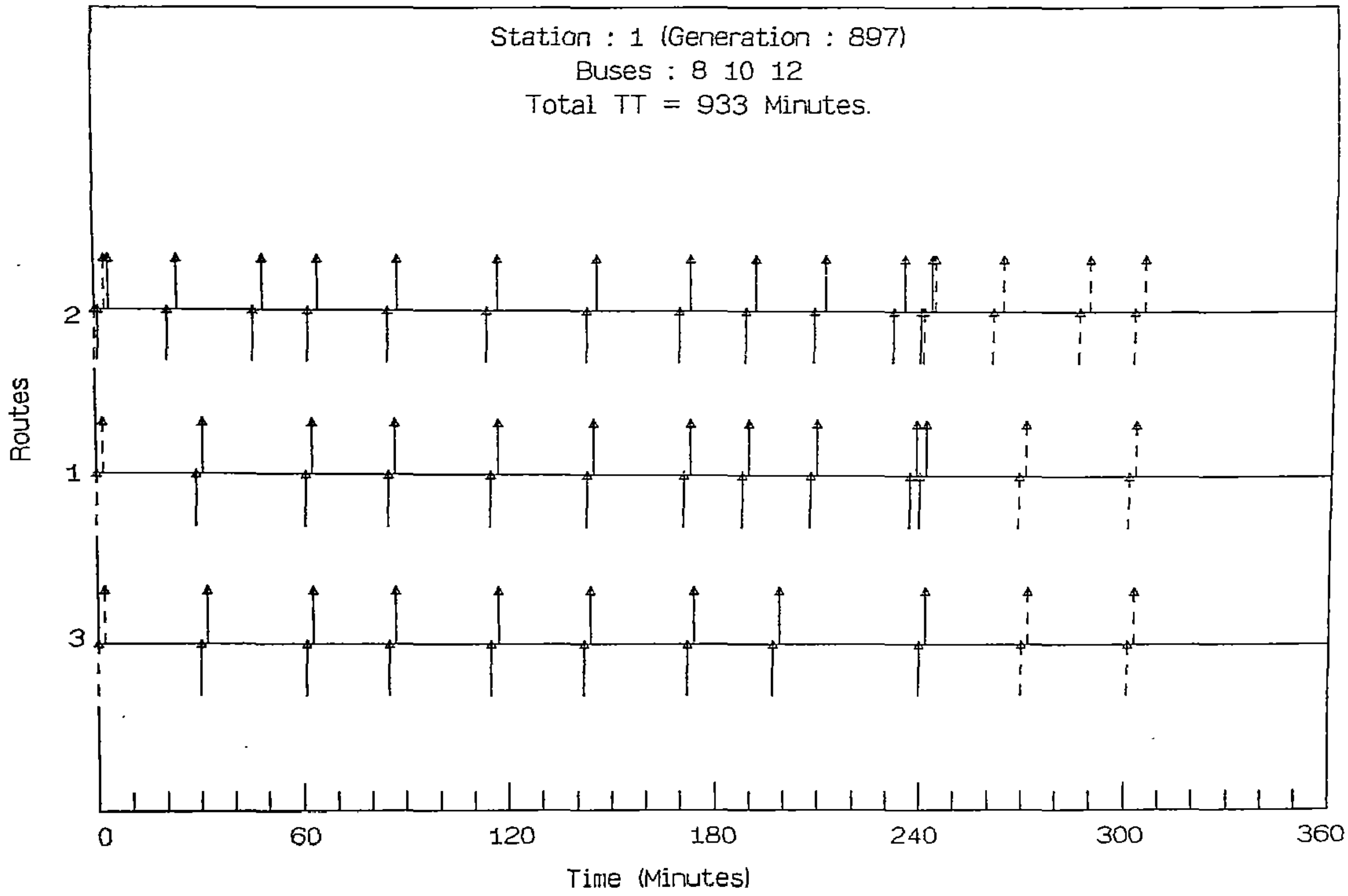


Figure 6.21(*a*): Optimal schedule for TT consideration

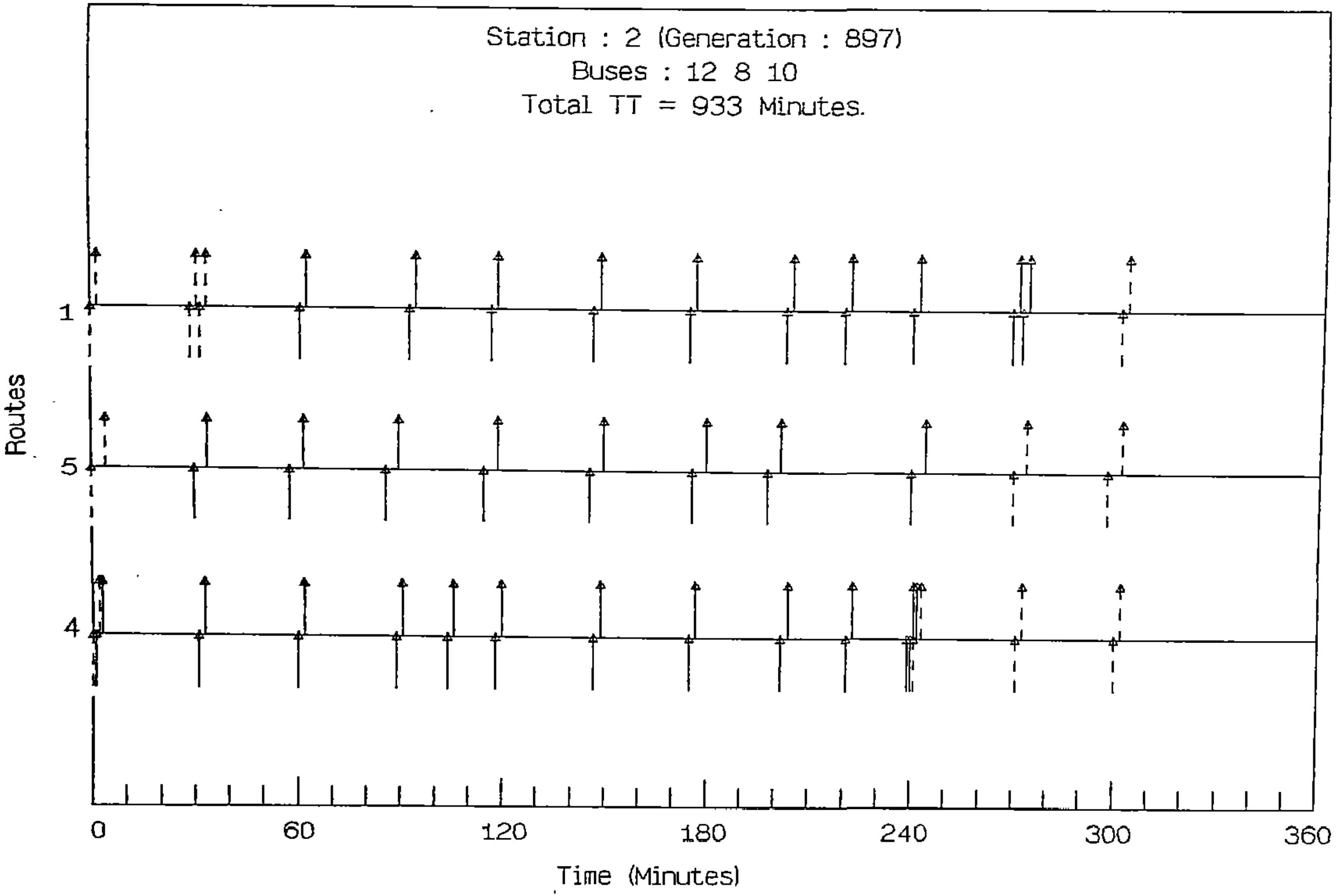


Figure 6.21(b): Optimal schedule for TT consideration

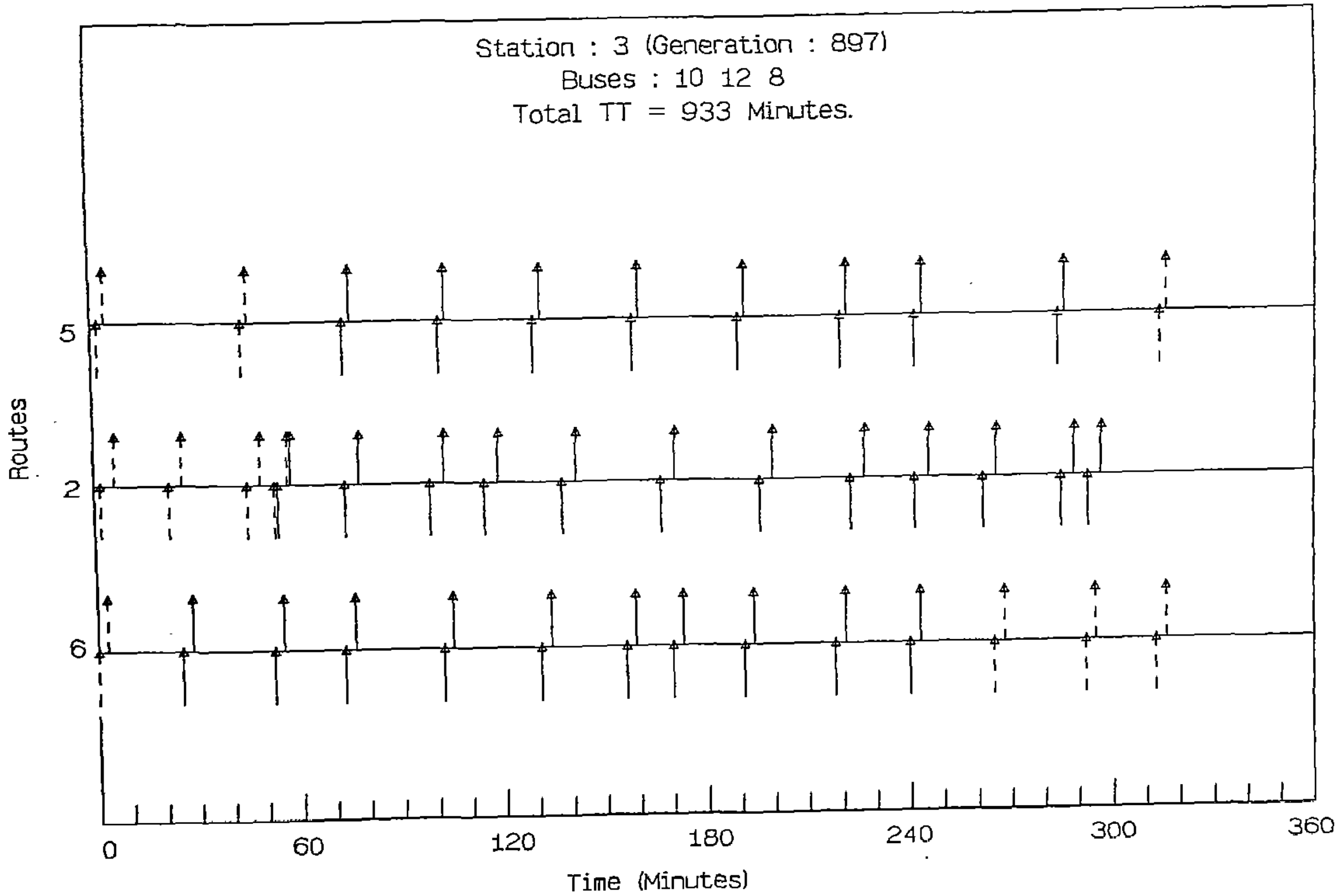


Figure 6.21(c): Optimal schedule for TT consideration

total TT of 568 minutes with equal resource on each route which is obvious. As seen from the figures, some headways are very small as the buses on the routes with large fleet size try to align with the last buses on routes with small fleet size at a station for minimum transfer time. This solution is one of the infinite solutions for only TT consideration.

Case - II

In this case, only IWT with uniform arrival pattern of passengers is considered. The optimal schedule is shown in the Figures 6.22(a), 6.22(b) and 6.22(c). As seen from figures, on all the routes at all the stations, the headways are uniform which is the optimal solution for IWT consideration. The total IWT is coming out to be 15576 minutes.

Case - III

In this case, both TT and IWT is taken into account. The schedule is shown in the Figures 6.23(a), 6.23(b) and 6.23(c). As there are unequal number of buses on each of the routes at a station and TT is also considered, some of the headways are different from uniform headway (which is the optimal case for IWT consideration), so that transfer time is also reduced. In this case transfer time is 1258 minutes which is more than that for only TT consideration, as IWT is also taken into account. The total waiting time obtained in this case is 17010 minutes.

Case - IV

The optimal schedule for IWT consideration with non-uniform arrival pattern of passengers is shown in the Figures 6.24(a), 6.24(b) and 6.24(c). The total TWT obtained is 15186 minutes. As more number of passengers come in between 120 to 240 minutes, more buses arrive in that period for overall waiting time for passengers to be minimum.

Case - V

The optimal schedule for TWT consideration with non-uniform arrival pattern of passengers is shown in the Figures 6.25(a), 6.25(b) and 6.25(c). In this case, it is not possible to predict any intuitive optimal schedule due to

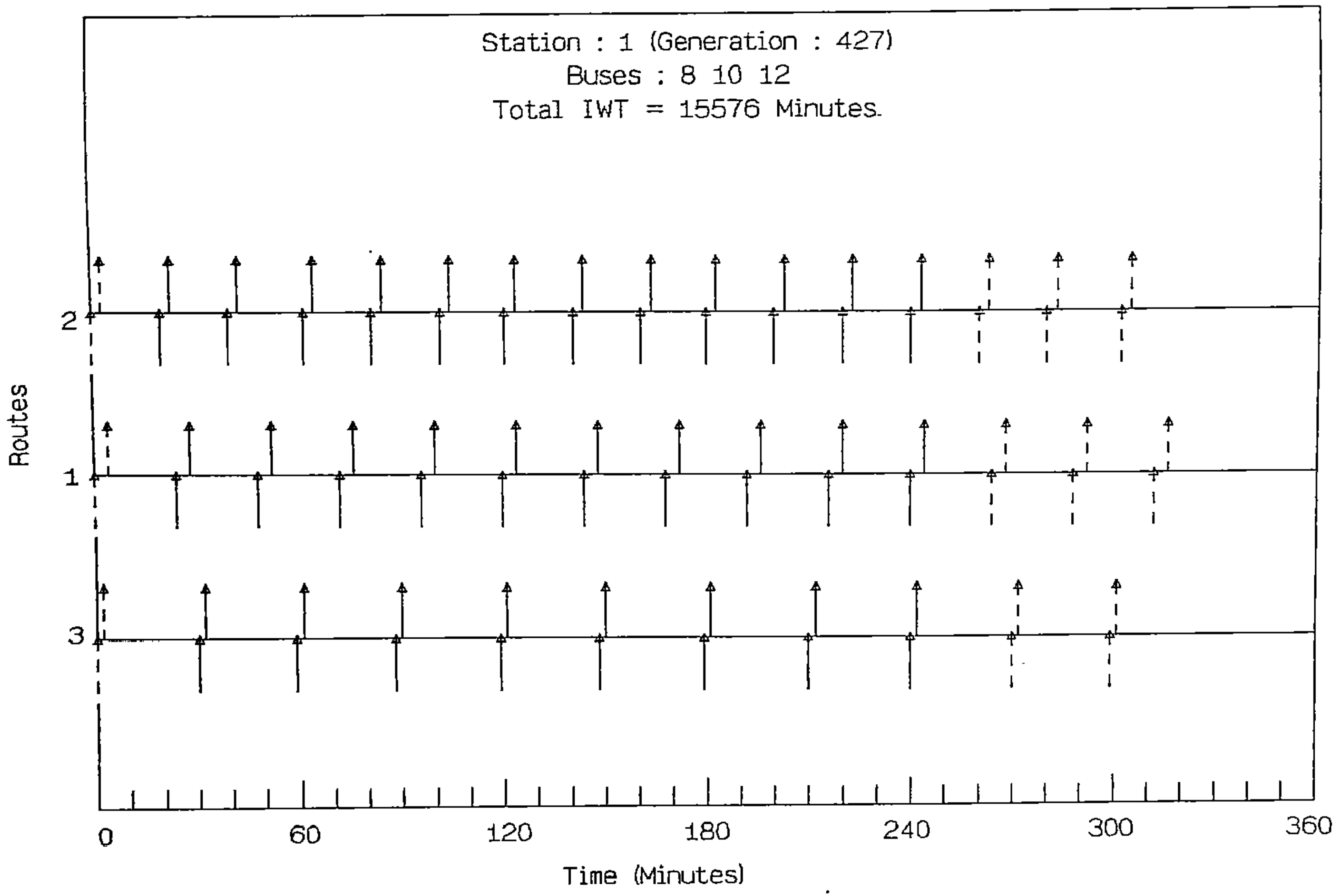


Figure 6.22(a): Optimal schedule for IWT consideration (Uniform arrival pattern)

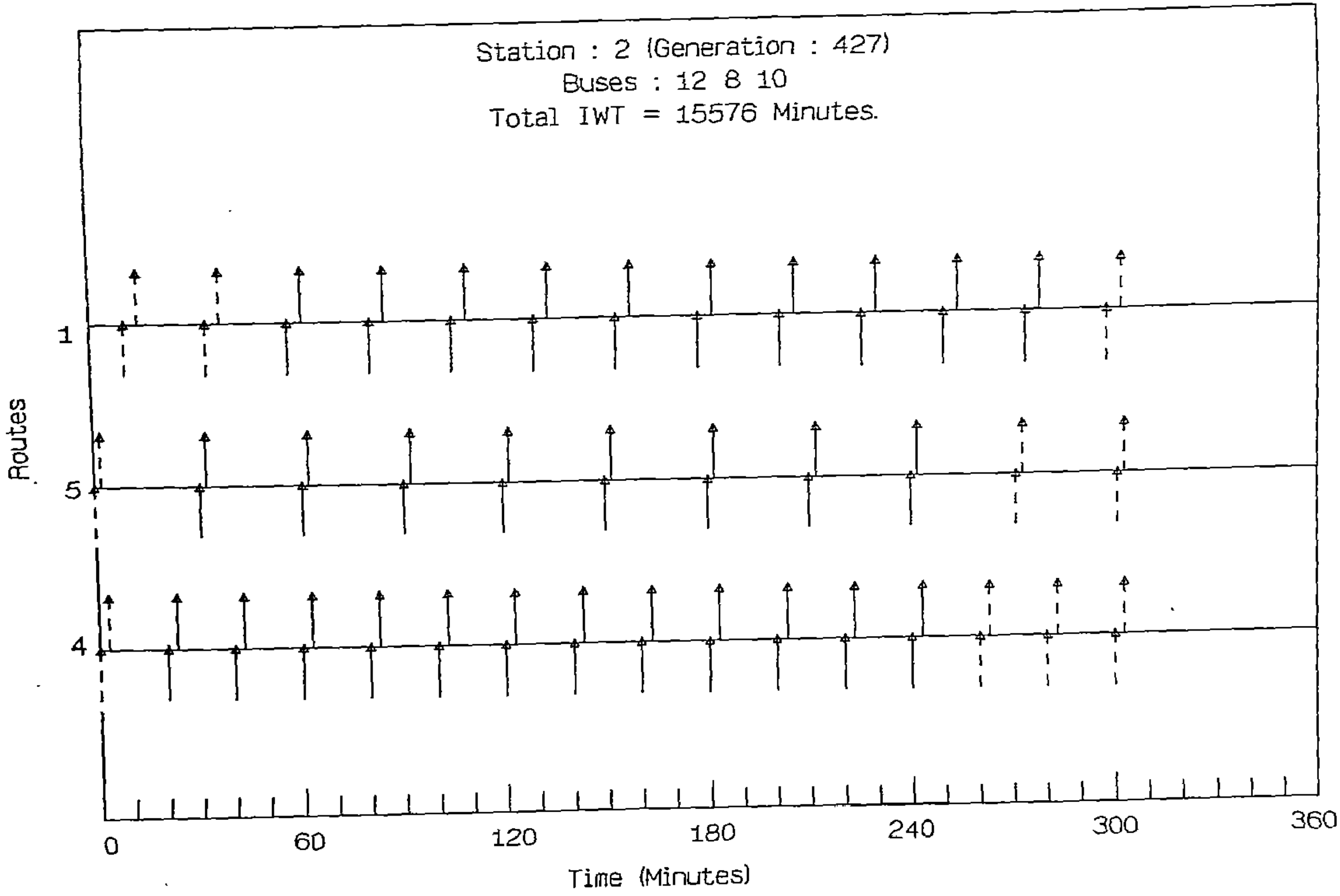


Figure 6.22(b): Optimal schedule for IWT consideration (Uniform arrival pattern)

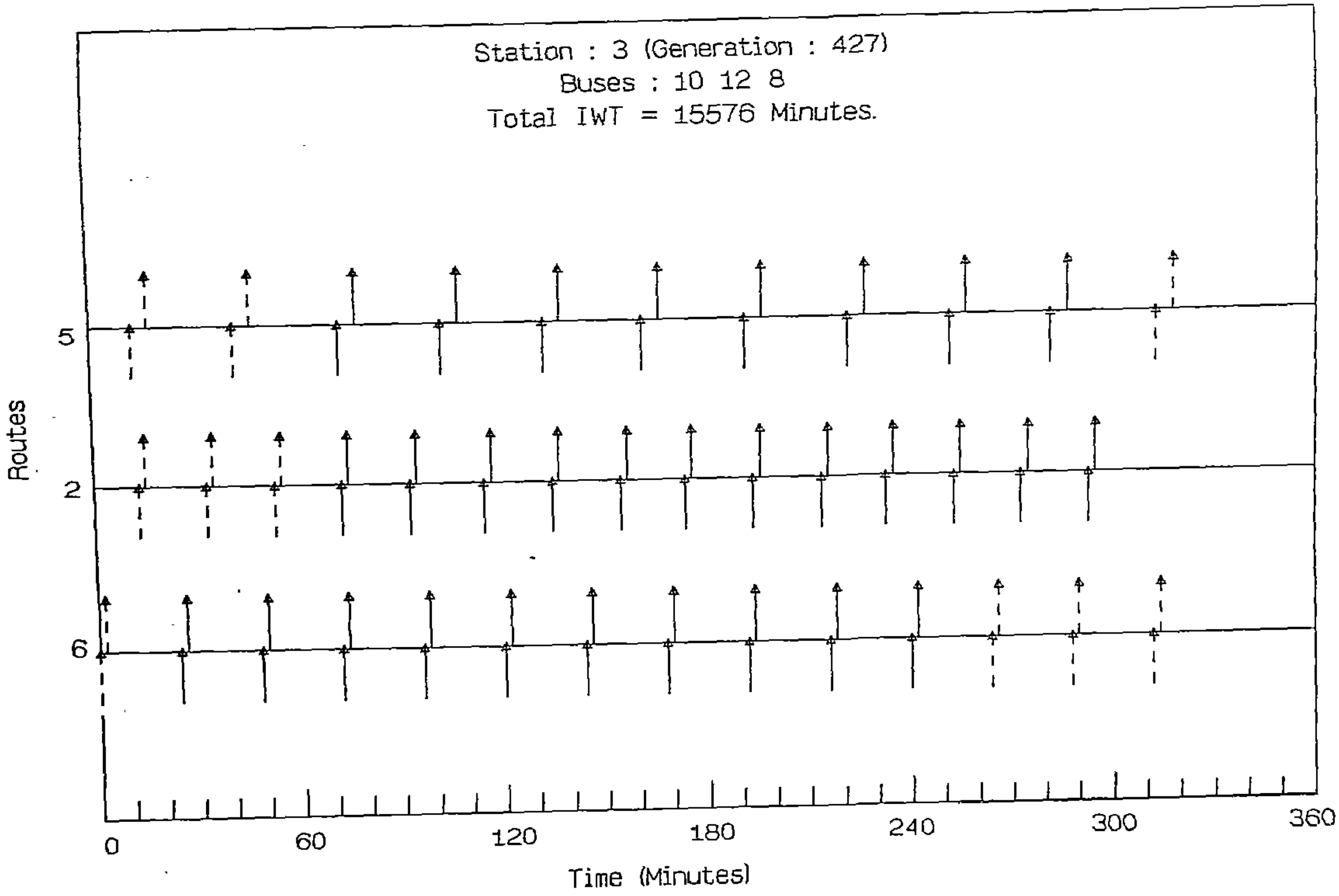


Figure 6.22(c): Optimal schedule for IWT consideration (Uniform arrival pattern)

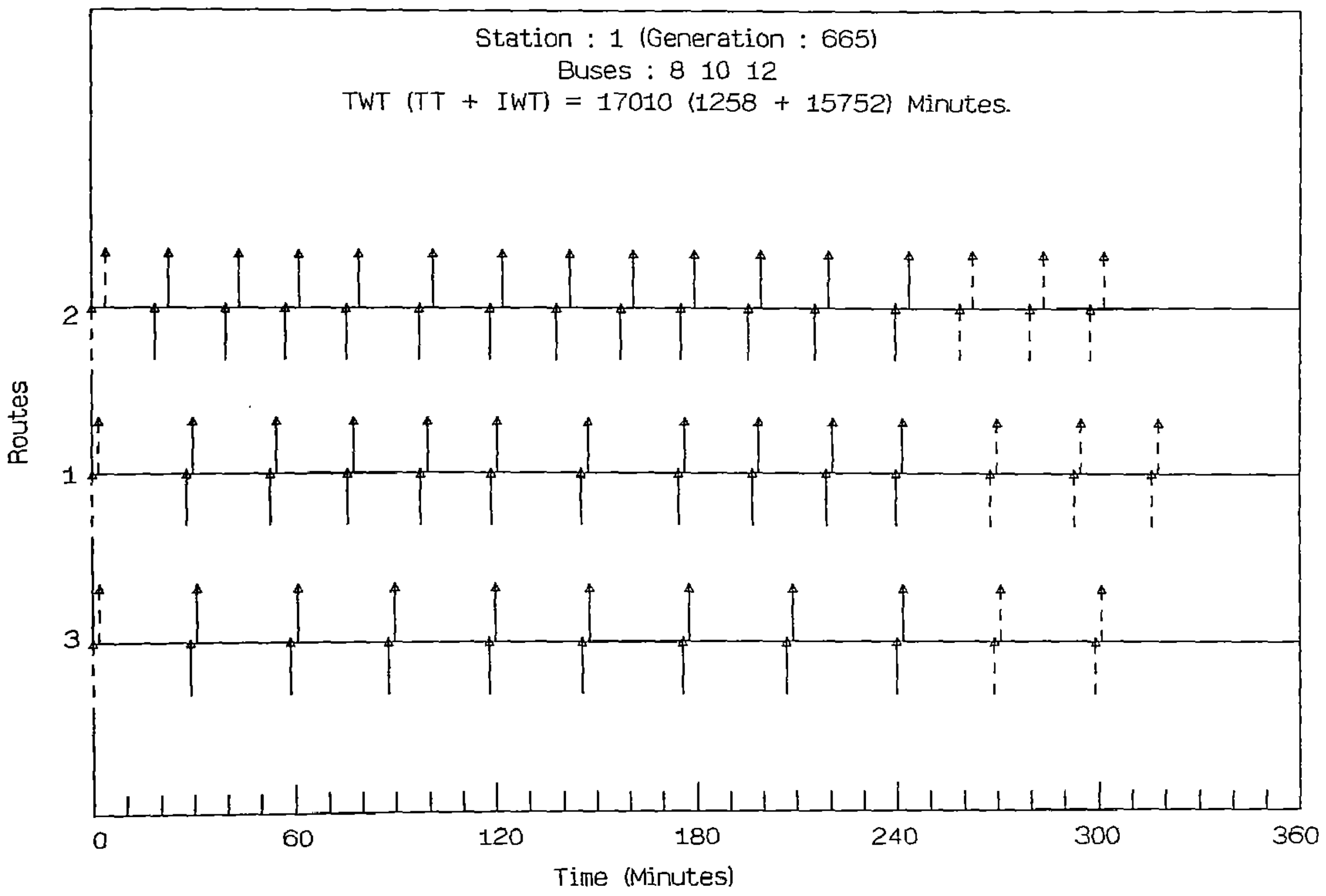


Figure 6.23(a): Optimal schedule for TWT consideration (Uniform arrival pattern)

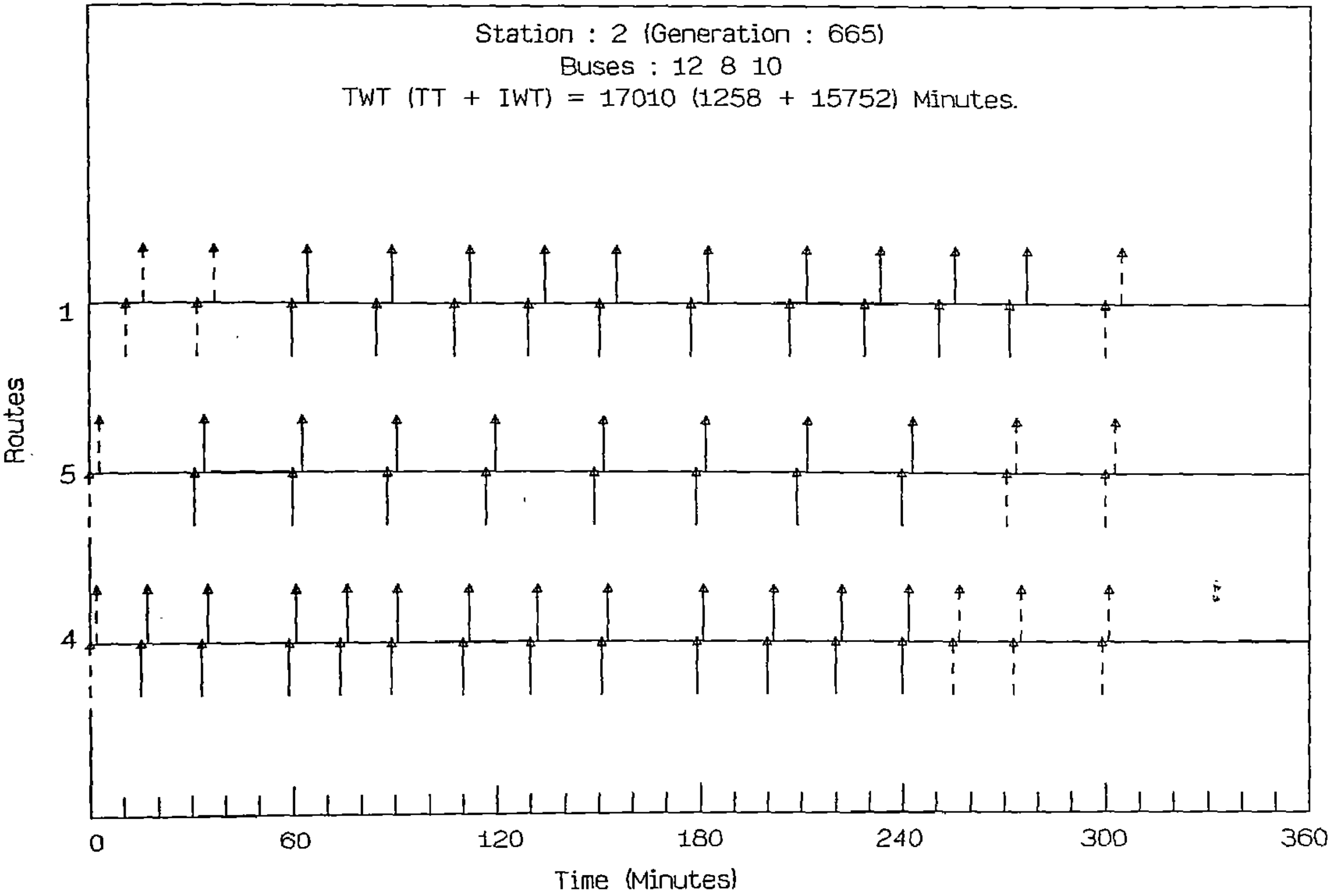


Figure 6.23(*b*): Optimal schedule for TWT consideration (Uniform arrival pattern)

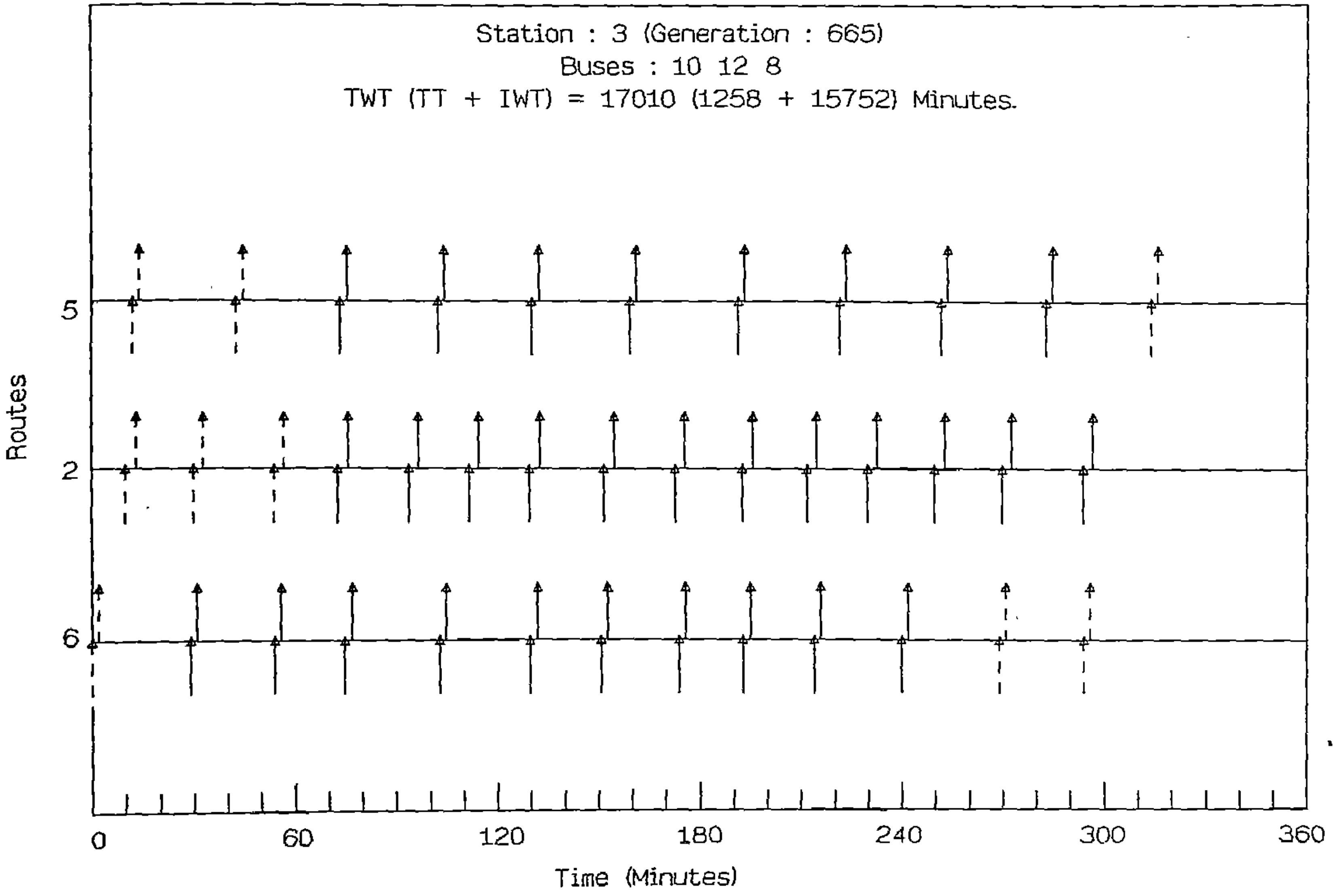


Figure 6.23(*c*): Optimal schedule for TWT consideration (Uniform arrival pattern)

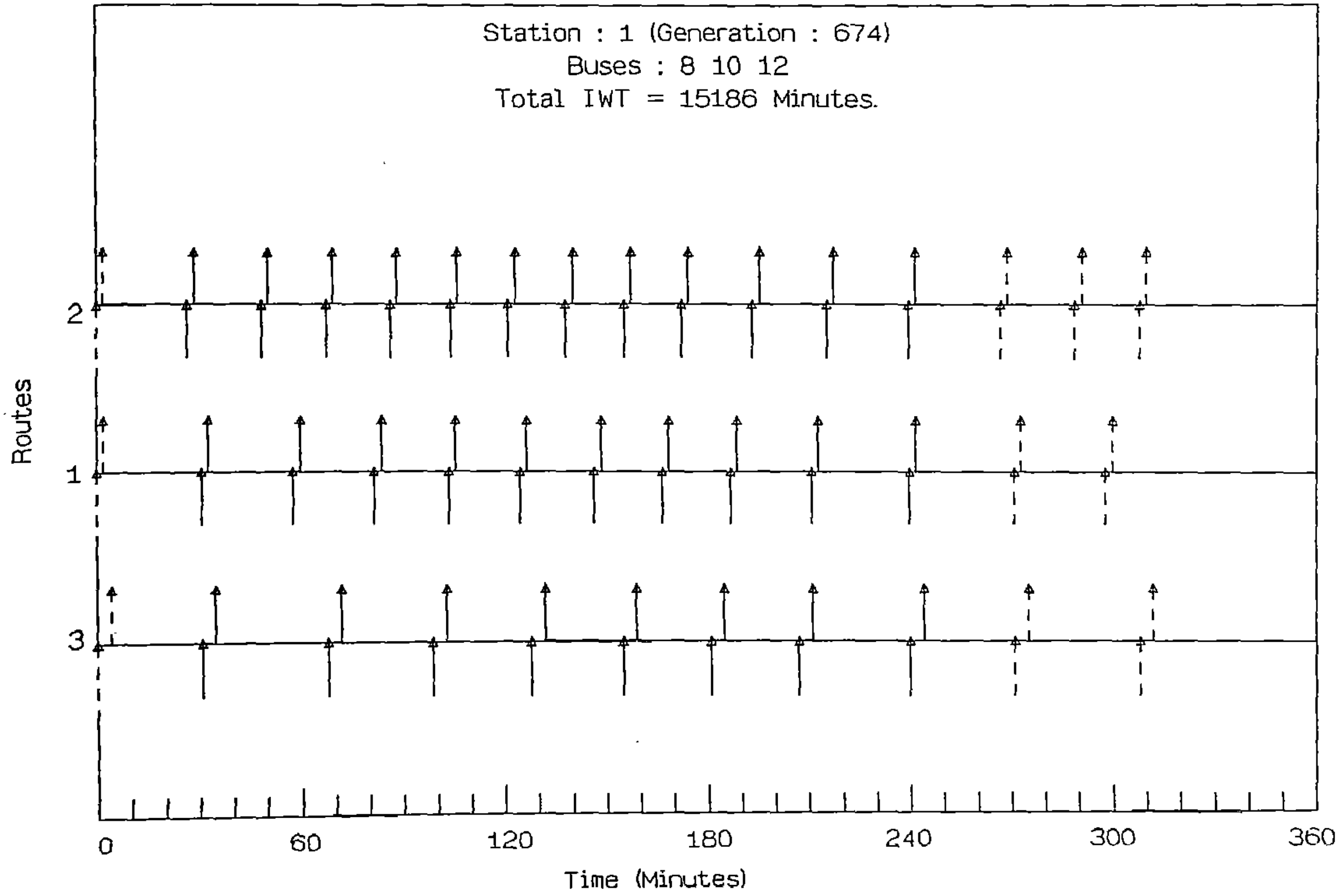


Figure 6.24(a): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

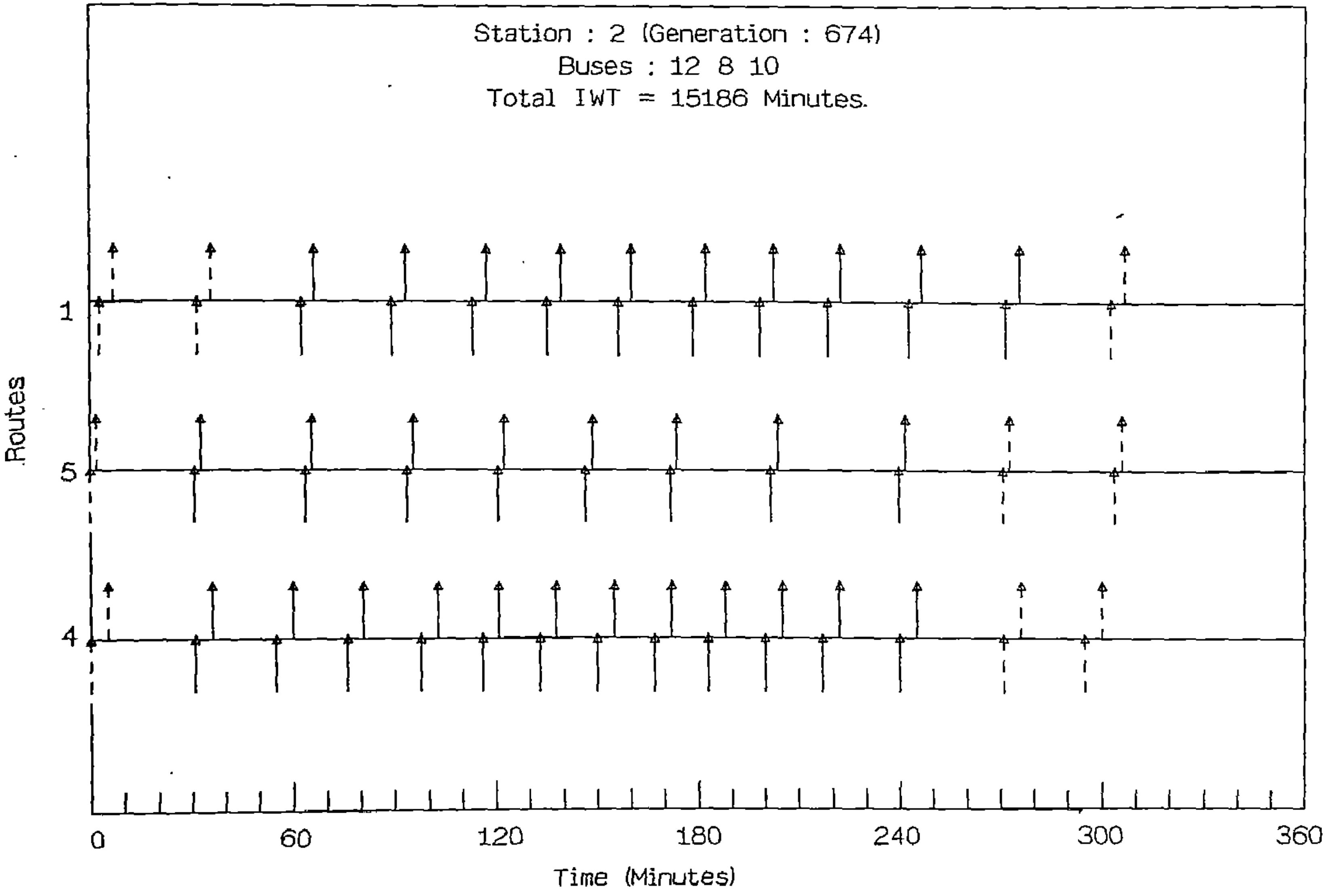


Figure 6.24(*b*): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

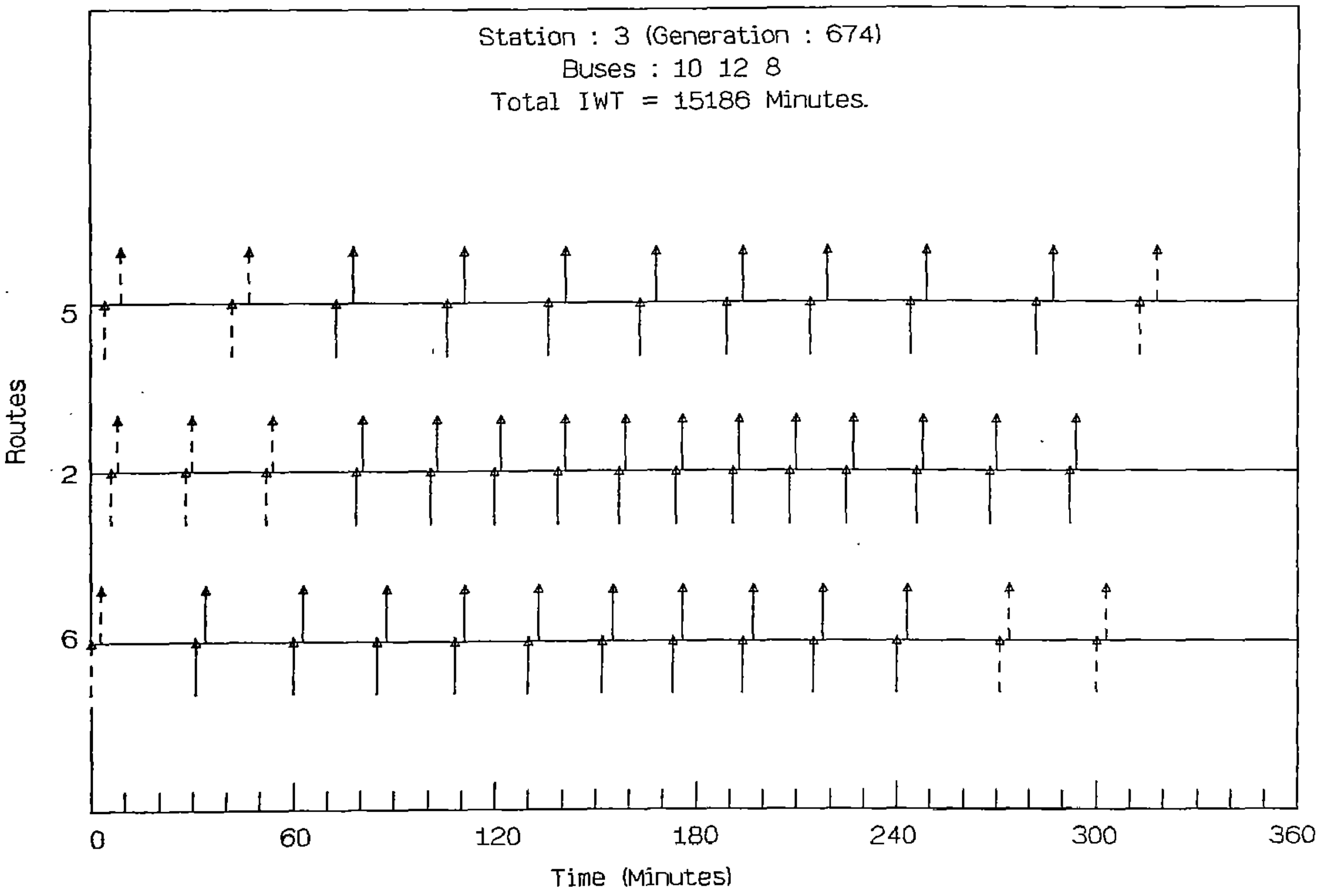


Figure 6.24(c): Optimal schedule for IWT consideration (Non-Uniform arrival pattern)

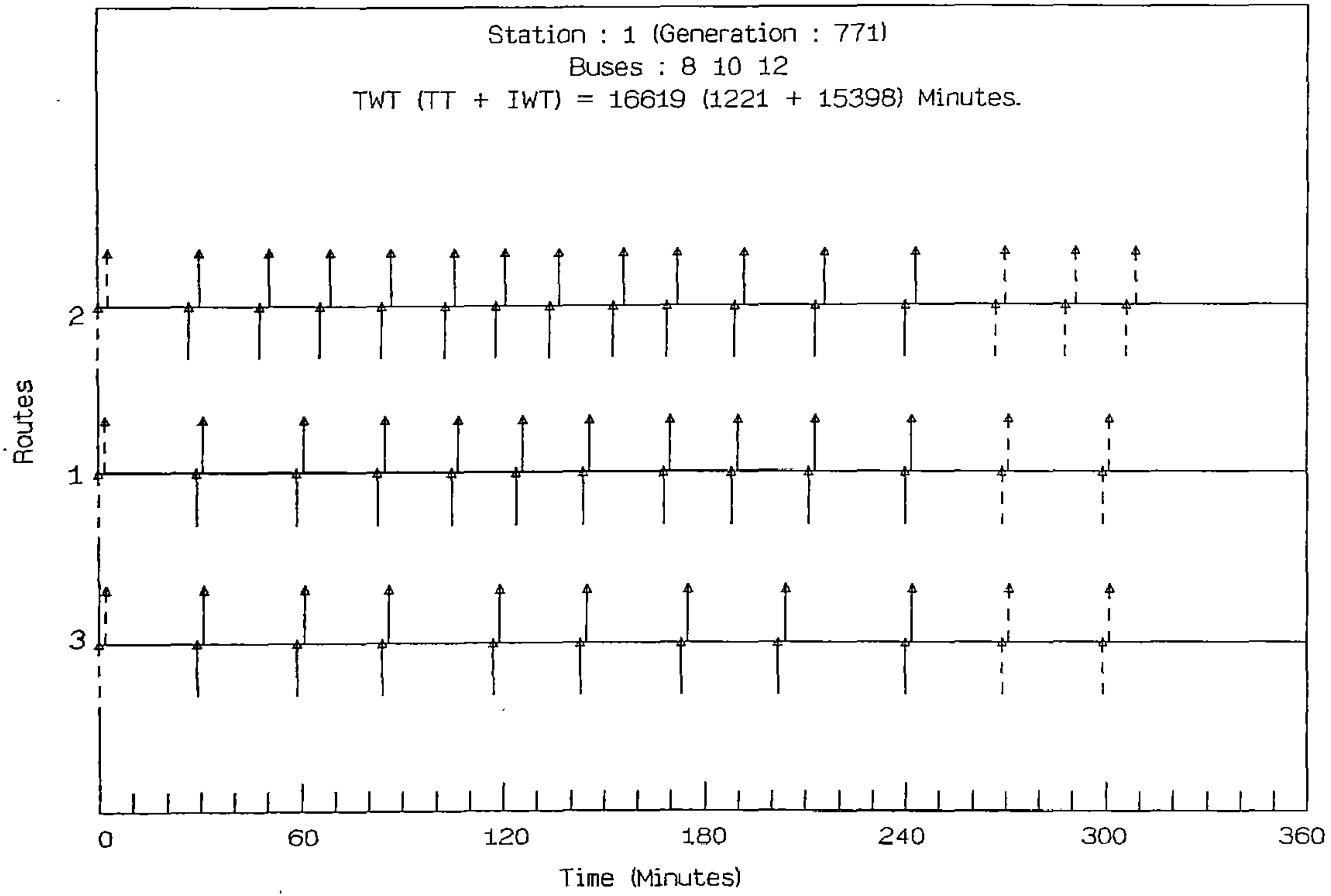


Figure 6.25(a): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

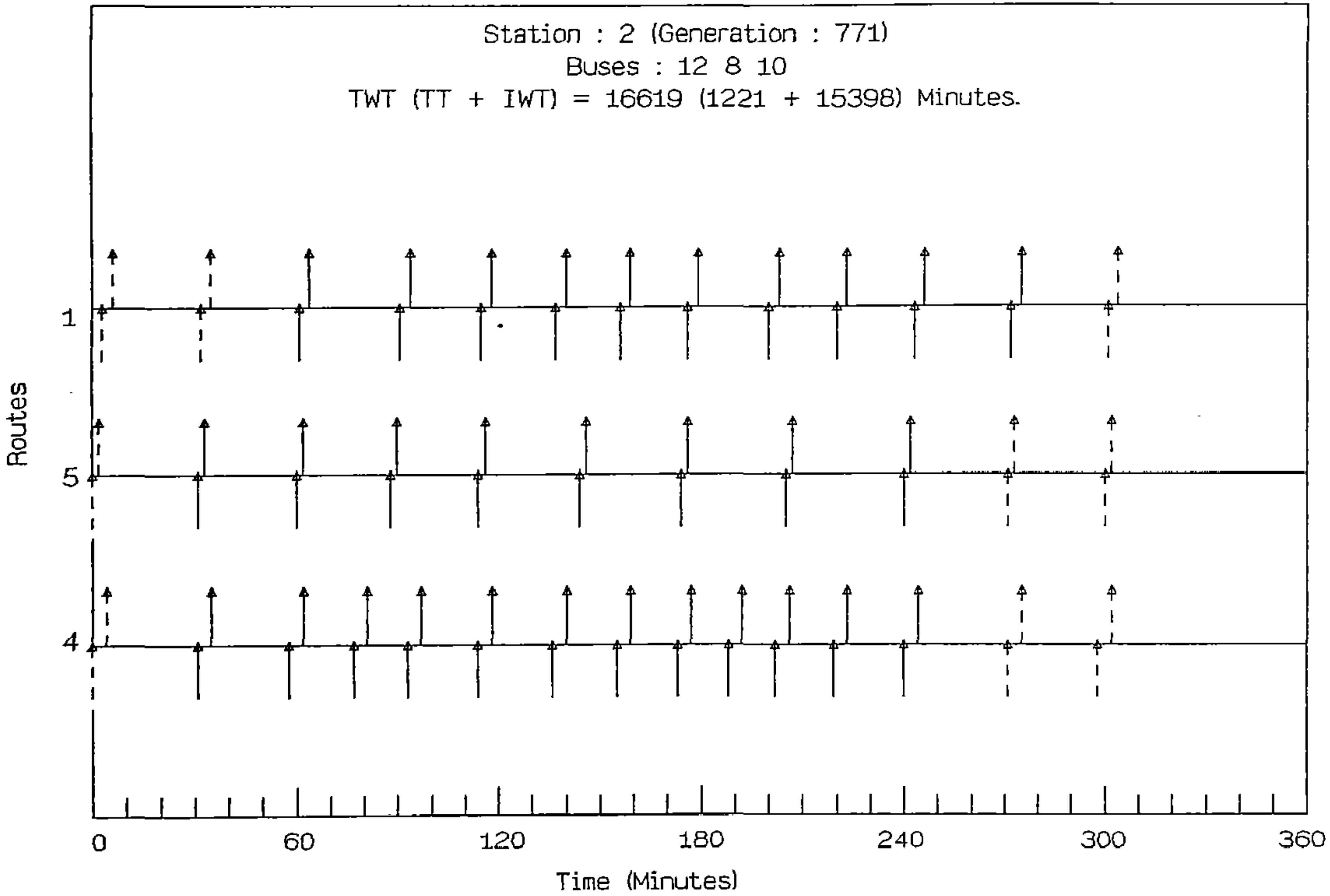


Figure 6.25(b): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

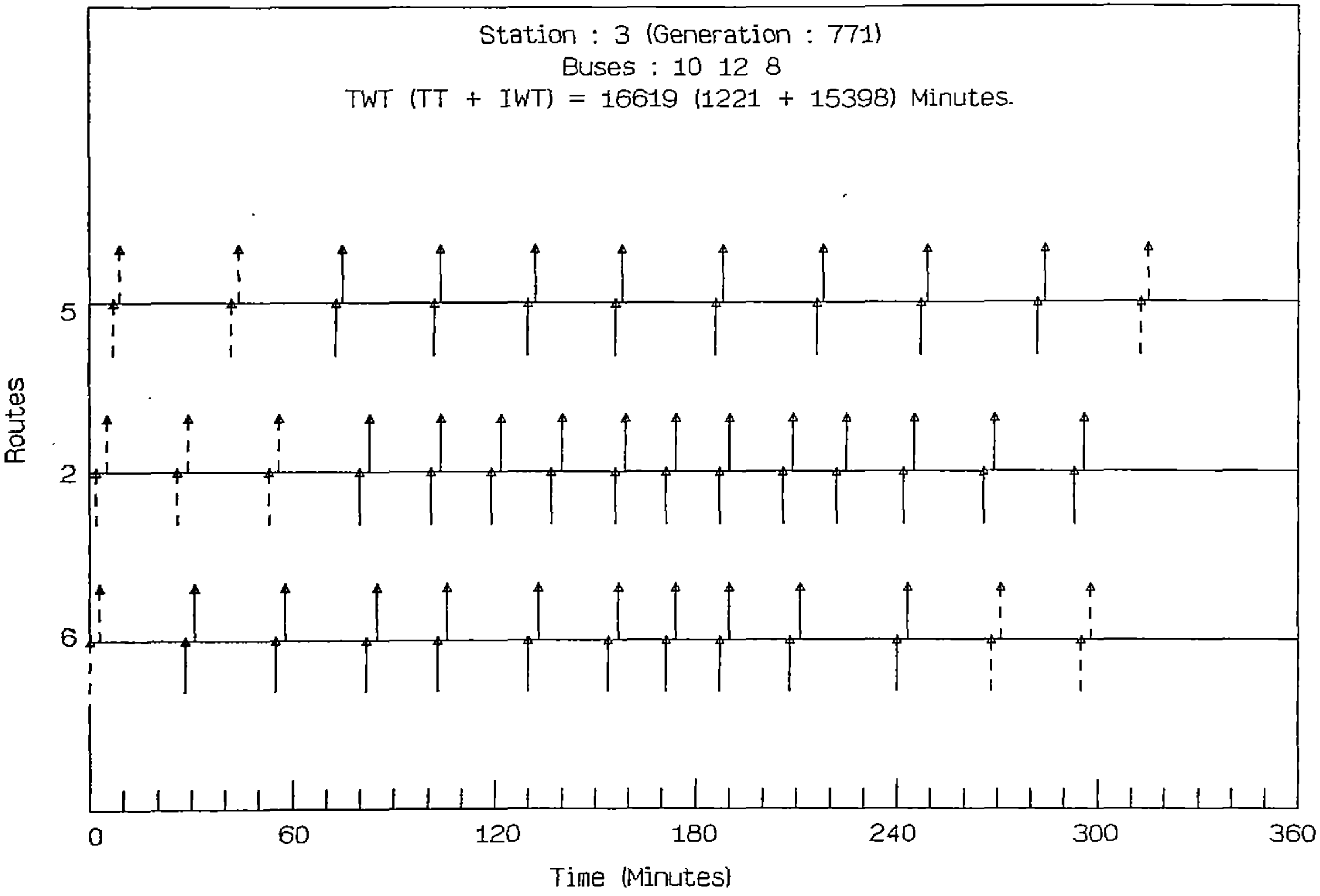


Figure 6.25(c): Optimal schedule for TWT consideration (Non-Uniform arrival pattern)

the introduction of TT term in the objective function. The TWT obtained in this case is 16619 minutes.

Chapter 7

Conclusions

In this study, an algorithm for scheduling of bus transit systems in a network of routes was developed. At each of the stations in a network, optimal schedules were found by minimizing the total waiting time of passengers which is one of the good measures of level of service. The objective function in the scheduling problem can be minimized by various optimization techniques, but each one has its advantages and disadvantages.

First, the mathematical programming formulation of the scheduling problem was presented. In the formulation, the objective function as well as some of the constraints are non-linear. As the mathematical programming problem is large, has both real and integer variables, and is non-linear in terms of decision variables, it is very difficult for classical programming techniques to solve the problem. Earlier study [4] showed that even after linearization also, the complexity involved in the problem is very large. Traditional methods proved futile in developing an optimal schedule even for simple scheduling problem with only one station. Our study involves multiple transfer stations which increases complexity further and so it cannot be solved with traditional methods.

In this study, *Genetic Algorithms*, an evolutionary optimization technique was used to solve the scheduling problem. The main advantage of using GAs is that the problem can be reformulated in a manner which is much simpler in computational complexity than the classical formulation. The modified formulation of the scheduling problem using procedure based function declarations and coding of variables showed how the complexity reduced in the problem; procedure based function declarations eliminate the constraints containing integer variables and coding of variables eliminate those constraints which provide bounds on different variables. Results from test problems demonstrate that the GAs are able to find optimal schedules with minimum computational resource.

Two simple networks were chosen as test networks and the schedules obtained for them were presented in the previous chapter. The schedules were developed for various cases, by minimizing each of the components in the objective function as well as both, with equal and unequal resource on each of the routes at a station. The envelope function defining the arrival pattern of passengers is assumed uniform as well as varying with time. The schedules obtained for TT/IWT with equal resource on each route (for which optimal solutions are known) were used to show that we get optimal solutions for the various cases. From the results, it can be infered that the adopted procedure of scheduling works effectively in attaining optimal schedules.

References

[1] Bodin L. and Golden B. (1981), "Classification in vehicle routing and scheduling", Networks, Vol.11, pp.97-108.

[2] Bookbinder J.H. and Desilets A. (1992), "Transfer optimization in a transit network", Transportation science, Vol.26, No.2, pp.106-118 .

[3] Ceder A. (1986), "Methods for creating bus time tables", Transportation research, Vol.21A, No.1, pp.55-83.

[4] Chakroborty P., Deb K. and Subrahmanyam P.S. (1995), "Optimal scheduling of urban transit systems using Genetic algorithms", Journal of Transportation Engineering (ASCE), Vol.121, No.6, pp.544-553.

[5] Chang S.K. and Schonfeld P.M. (1991), "Multiple period optimization of bus transit systems", Transportation research, Vol.25B, No.6, pp.453-478.

[6] Deb K. (1993), "Genetic algorithms in Engineering design optimization", In J.N.Reddy et al. (Eds), Proceedings of the Advanced study institute on computational methods for engineering analysis and design, pp.12.1-12.25.

[7] Friedman M. (1976), "A mathematical programming model for optimal scheduling of buses departures under deterministic conditions", Transportation research, Vol.10. No.2, pp.83-90.

[8] Glover F. (1975), "Improved linear integer programming formulation of non-linear integer problems", Management science, Vol.22, pp.455-460.

[9] Goldberg D.E. (1989), "Genetic algorithms in search, optimization and machine learning", MA, Addison-Wesley.

[10] Goldberg D.E. and Deb K. (1991), "A comparative analysis of selection schemes used in Genetic algorithms", Foundations of Genetic algorithms, Edited by G.J.E. Rawing, pp.69-93.

[11] Horowitz A.J. (1981), "Subjective value of time in bus transit travel", Transportation 10, pp.149-164.

[12] Hutchinson T.P. and Jolliffe J.K. (1975), "A behavioural explanation of the association between bus and passenger arrivals at a bus stop", Transportation science, Vol.9, No.3, pp.248-282.

[13] Kikuchi S. and Parameswaran J. (1993), "Solving a schedule coordination problem using fuzzy control technique", Presented at Intelligent scheduling systems symposium at the ORSA/TIMS conference, San Francisco, CA.

[14] Lenstra J.K. and Rinnooy Kan A.H.G. (1981), "Complexity of vehicle routing and scheduling problems", Networks, Vol. 11, pp.221-227.

[15] Rekailtis G.V., Ravindran A. and Ragsdell K.M. (1983), "Engineering optimization - Methods and applications", New york, Wiley.

[16] Subrahmanyam P.S. (1995) "Transit system scheduling : Use of Genetic algorithms", M.Tech thesis submitted to the department of civil engineering, IIT Kanpur.

121391

CE-1995-M-SRI-NET

A121391